



MEDICAL COUNCIL
OF CANADA

LE CONSEIL MÉDICAL
DU CANADA

**Medical Council
of Canada
Qualifying
Examination
(MCCQE) Part I**

**MCCQE PART I
ANNUAL
TECHNICAL
REPORT**

2020

JUNE 2020 — JAN. 2021

Table of contents

Preface	4
1. Overview of the MCCQE Part I	4
2. Exam development.....	6
2.1 Exam blueprint	6
2.2 Exam specifications.....	8
2.2.1 <i>Content specifications</i>	9
2.2.2 <i>Psychometric specifications</i>	10
2.3 Item development.....	11
2.3.1 <i>Test committees</i>	11
2.3.2 <i>Clinical Decision-Making questions</i>	12
2.4 Test assembly	14
3. Exam administration.....	16
3.1 Exam delivery	16
3.2 Exam security	16
3.3 Exam preparation	18
3.4 Quality assurance.....	18
3.5 Release of results.....	18
4. Validity	19
4.1 The argument-based approach to validation.....	19
5. Psychometric analyses.....	26
5.1 Item analysis: classical test theory and item response theory	26
5.2 Item calibration	28
5.3 Estimating candidate ability.....	29
5.4 Standard setting and scaling	30
5.5 Score reporting.....	31
6. Exam results.....	32
6.1 Candidate cohorts	32
6.2 Overall exam results.....	33
6.3 Reliability of exam scores and classification decisions	35
6.4 Pass or fail decision accuracy and consistency	37
6.5 Domain subscore profile.....	38
6.6 Historical pass rates	41
6.7 Candidate survey	41
7. References	45
APPENDIX A: MCCQE Part I Statement of Results sample.....	47
APPENDIX B: MCCQE Part I Supplemental Information Report sample	48
APPENDIX C: Internal structure of the MCCQE Part I	51

List of Tables and Figures

Table 1: Blueprint for the MCCQE Part I.....	7
Table 2: MCCQE Part I exam specification.....	8
Figure 1: Target test information function.....	10
Figure 2: Automated test assembly procedure.....	15
Figure 3: Key elements in Kane’s argument-based approach to validation: Inferences from observation to decision.....	20
Table 3: Level of inference – Scoring.....	21
Table 4: Level of inference – Generalization.....	23
Table 5: Level of inference – Extrapolation.....	24
Table 6: Level of inference – Decisions.....	25
Table 7: MCCQE Part I group composition, 2020–2021.....	32
Table 8: MCCQE Part I results, 2020–2021.....	33
Figure 4: MCCQE Part I total score distributions, 2020–2021.....	34
Table 9: Reliability estimates, standard errors of measurement, decision consistency and decision accuracy indices for each MCCQE Part I session, 2020–2021.....	38
Figure 9: Domain subscore for the MCCQE Part I, June 2020 session.....	39
Figure 10: Domain subscore for the MCCQE Part I, July 2020 session.....	39
Figure 11: Domain subscore for the MCCQE Part I, October 2020 session.....	40
Figure 12: Domain subscore for the MCCQE Part I, January 2021 session.....	40
Table 10: MCCQE Part I pass rates, April 2017 to January 2021.....	41
Table 11: MCCQE Part I post-MCQ candidate survey results, 2020–2021 Percentages for ratings.....	42
Table 12: MCCQE Part I post-CDM candidate survey results, 2020–2021 Percentages for ratings.....	43
Table 13: Correlation matrix among subscores in the four domains of Dimensions of Care and total scores.....	51
Table 14: Correlation matrix among subscores in the four domains of Physician Activities and total scores.....	52
Table 15: Correlation matrix among subscores in Physician Activities and in Dimensions of Care.....	52

Preface

This report summarizes the fundamental psychometric characteristics, test development, test publishing, and test administration activities of the MCCQE Part I. Candidate performance data on the exam in June 2020 to January 2021¹ are also presented. Sections 1 to 5 describe the exam's purpose, format, content development, administration, scoring and score reporting. These sections also provide evidence in support of score interpretation, reliability and errors of measurement, and other psychometric characteristics. Section 6 summarizes candidate performances for the four sessions in 2020–2021 and includes historical data for reference purposes. The report serves as technical documentation and reference materials for members of the Central Examination Committee (CEC), test committee members, Medical Council of Canada (MCC) staff, members of the Council, stakeholders, and members of the public.

1. Overview of the MCCQE Part I

The MCCQE Part I is a summative exam that assesses the critical medical knowledge and clinical decision-making (CDM) ability of a candidate at a level expected of a medical student who is completing their medical degree in Canada. The exam is based on the *MCC Objectives*, which are organized under the *CanMEDS roles* (Frank, Snell & Sherbino, 2015). Candidates graduating and successfully completing the MCCQE Part I typically enter supervised practice. In addition to the formal accreditation processes of the undergraduate and postgraduate education programs, the MCCQE Part I is the only national standard for medical schools across Canada and is administered at the end of medical school.

The MCCQE Part I is a one-day computer-based test. Candidates have four hours in the morning session to complete 210 *multiple-choice questions* (MCQs) and three and a half hours in the afternoon session for the *CDM component*, which consists of 38 cases with short-menu and short-answer write-in questions.

¹ 2020–2021 includes data from the June, July and October 2020 and January 2021 exam windows.

The 2009 strategic review

In fall 2009, MCC undertook a strategic review of its assessment processes that focused on their purposes, objectives, structure and alignment with MCC's major stakeholder requirements. The review addressed current trends in medical education, regulation and assessment. It also considered the role and function of MCC's exams in meeting the current and future needs of medical regulatory authorities, the public and other stakeholders. It also focused on validating and updating the Blueprint for both components of the MCCQE Part I.

The 2014 Blueprint project

MCC undertook the *Blueprint project* to review and establish an evidence-based approach for identifying the competencies that physicians will be expected to demonstrate and be assessed on at two decision points: (1) entry into residency and (2) entry into independent practice.

The Blueprint project's purpose was to ensure that critical core competencies, knowledge, skills and behaviours for safe and effective patient care in Canada are being appropriately assessed for the two decision points. A new Blueprint for MCCQE Part I was approved by the Council in 2014 (see Exam Blueprint, section 2.1).

The emergence of COVID-19

The MCC was able to adapt quickly during the COVID-19 pandemic because it had partnered with Prometric in 2019, which already had the solution to administer the exam both onsite in test centres and remotely using live online proctors. In 2020, MCC introduced remote proctoring as a new test modality and alternative for delivering the MCCQE Part I. Remote proctoring supports MCC's assessment of core competencies of physicians before specialty training or certification. It allows MCC to continue administering exams when physical distancing is required and test centres may not be available.

The Central Examination Committee (CEC)

The CEC is responsible for overseeing the MCCQE Part I, including the:

- Exam Blueprint
- Exam specifications and constraints
- Development of the exam
- Maintenance of exam content
- Ruling on candidate special cases and reconsiderations
- Approval of results

2. Exam development

In this section, we describe the exam Blueprint, exam specifications, item development and test assembly.

2.1 EXAM BLUEPRINT

Exam development begins with the exam Blueprint, which was approved by Council in 2014. The content specifications for the MCCQE Part I were approved by the CEC in 2016. The Blueprint addresses candidates' performance across two broad categories: Dimensions of Care and Physician Activities. There are four domains of care under each of these categories.

1. **Dimensions of Care** reflects the focus of care for the patient, family, community and/or population. Its four domains are as follows:
 - a. **Health promotion and illness prevention:** the process of enabling people to increase control over their health and its determinants, and thereby improve their health. Illness Prevention covers measures not only to prevent the occurrence of illness, such as risk factor reduction, but also to arrest its progress and reduce its consequences once established. This includes but is not limited to screening, periodic health exam, health maintenance, patient education and advocacy, and community and population health.
 - b. **Acute:** brief episode of illness within the time span defined by initial presentation through to transition of care. This dimension includes but is not limited to urgent, emergent and life-threatening conditions, new conditions, and exacerbation of underlying conditions.
 - c. **Chronic:** illness of long duration that includes but is not limited to illnesses with slow progression.
 - d. **Psychosocial aspects:** presentations rooted in the social and psychological determinants of health and how these can impact well-being or illness. The determinants include but are not limited to life challenges, income, culture, and the impact of the patient's social and physical environment.
2. **Physician Activities** reflects the scope of practice and behaviours of a physician practising in Canada and has four domains:
 - a. **Assessment and diagnosis:** exploration of illness and disease using clinical judgment to gather, interpret and synthesize relevant information that includes but is not limited to history taking, physical examination and investigation.

- b. **Management:** process that includes but is not limited to generating, planning, organizing safe and effective care in collaboration with patients, families, communities, populations and other professionals (e.g., finding common ground, agreeing on problems and goals of care, time and resource management, roles to arrive at mutual decisions for treatment, working in teams).
- c. **Communication:** interactions with patients, families, caregivers, other professionals, communities and populations. Elements include but are not limited to relationship development, intra- and inter-professional collaborative care, education, verbal communication (e.g., using patient-centred interviews and active listening), non-verbal and written communication, obtaining informed consent and disclosure of patient safety incidents.
- d. **Professional behaviours:** attitudes, knowledge and skills related to clinical and/or medical administrative competence, communication, ethics, as well as societal and legal duties. The wise application of these behaviours demonstrates a commitment to excellence, respect, integrity, empathy, accountability and altruism within the Canadian health-care system. Professional behaviours also include but are not limited to self-awareness, reflection, lifelong learning, leadership, scholarly habits and physician health for sustainable practice.

Blueprint for the MCCQE Part I

Table 1 displays the Blueprint and associated content specifications (content weightings) for the MCCQE Part I. Both categories, Dimensions of Care and Physician Activities, have four domains, and each domain is assigned a specific content weighting on the exam.

Table 1: Blueprint for the MCCQE Part I

		Dimensions of care				
		Health Promotion & Illness Prevention	Acute	Chronic	Psychosocial Aspects	Row %
Physician activities	Assessment/ Diagnosis					45±5
	Management					35±5
	Communication					10±5
	Professional Behaviours					10±5
Column %		20±5	35±5	30±5	15±5	100

2.2 EXAM SPECIFICATIONS

MCC has developed content specifications that include certain constraints and psychometric specifications to test a broad sampling of topics and populations in medicine as outlined in the Blueprint. While the exam is divided into an MCQ component and a CDM component for delivery purposes, content and psychometric specifications are considered at the total test level.

Table 2 displays the exam specifications for the MCCQE Part I.

Table 2: MCCQE Part I exam specification

CATEGORY	DESCRIPTION
Complexity	Multiple morbidities
Age	<ul style="list-style-type: none"> • neonate • infant, child • adolescent • adult • adult women of childbearing age • frail elderly
Gender	Male, female, nonbinary
Special populations	Including but not limited to people who are: <ul style="list-style-type: none"> • Indigenous • LGBTQ2S+ • recent immigrants • living in rural areas • living with a disability • terminally ill • refugees • living with low incomes in inner city areas • living with substance use disorders • experiencing homelessness
Setting	Included but not limited to: <ul style="list-style-type: none"> • rural or remote settings • long-term-care institutions • home visits

2.2.1 Content specifications

The MCQ and CDM components of the MCCQE Part I are described in more detail below.

The MCQ component

The exam consists of 210 MCQs and includes pilot questions, also called pre-test questions, which are scored if they perform well psychometrically. The pilot questions are not identified as pilots in the exam. Each MCQ has a stem that includes a case description and three to five options (as of 2020), of which only one is the correct answer. Candidates may select only one option in the MCQs and points are not deducted for incorrect answers. The maximum time allotted for the MCQ component is four hours.

Certain questions will include visual material, such as a photograph, a radiograph, or an electrocardiogram. If relevant to a question, normal lab values are presented directly in the question stem.

The CDM component

The exam consists of 38 CDM cases and includes pilot questions, also called pre-test questions, which are scored if they perform well psychometrically. The pilot questions are not identified as pilots in the exam. Each question includes a stem that includes a case description followed by one or more options that assess problem-solving and decision-making skills in the resolution of a clinical case. Candidates may be asked to

- elicit clinical information
- order diagnostic procedures
- make diagnoses
- prescribe therapy

In total, candidates are presented with 69 to 75 questions related to the 38 CDM cases. Responses are either in a short-menu or short-answer write-in format.

Most questions explicitly state how many responses can be selected. Points are not deducted for incorrect answers. However, if a candidate exceeds the maximum number of allowable responses or selects a response that harms or endangers the patient, they receive a score of zero, even if they have also identified the correct answer. Some items ask candidates to “select as many as appropriate.” These question types require the candidate to narrow in on the investigation or diagnosis. Selecting too many responses may also result in the candidate receiving a zero, even if the correct answer is part of their answer choice. The maximum time allotted for the CDM component of the exam is three and a half hours.

Similar to the MCQ section, all cases and questions are typically presented in one continuous block of time. Certain questions will include visual material, such as a photograph, a radiograph, or an electrocardiogram. If relevant to the question, normal lab values are presented directly in the question stem or in the case.

2.2.2 Psychometric specifications

Psychometric specifications include the desired psychometric properties of the exam, which for the MCCQE Part I includes an overall target test information function (TIF) for each test form. The target TIF is used to balance multiple test forms and to ensure that precision of measurement across the ability scale is highly comparable from one test form to another. Figure 1 displays the target TIF. Test forms are assembled to control maximum information to be within ± 5 per cent of the target.

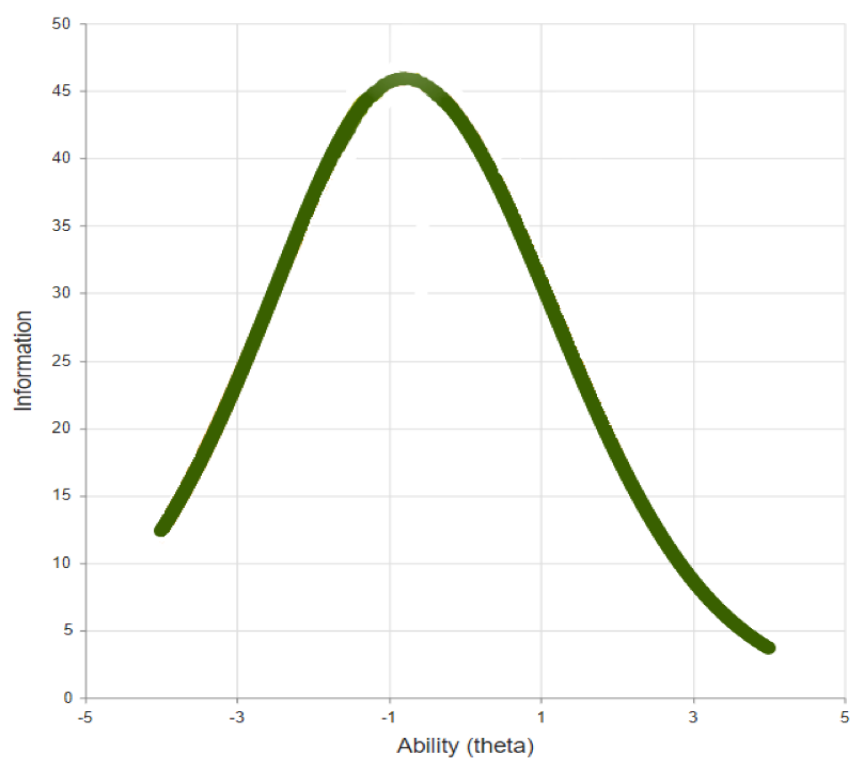


Figure 1: Target test information function

2.3 ITEM DEVELOPMENT

For MCQ content, six specialty test committees create and approve exam content. For CDM content, one multidisciplinary test committee develops exam content. In 2020, the MCCQE Part I test development officers (TDOs) held three content development workshops that served two purposes: (1) increase the quantity of new MCQ and CDM content and (2) educate item writers in the domain of MCQ and CDM content. The three workshops were as follows:

1. One focused on the development of MCQ content related to the opioid crisis in Canada (this work was subsidized through a grant from the Association of Faculties of Medicine in Canada).
2. A second focused on identifying gaps in the item bank.
3. A third was held to increase new CDM content.

MCC's medical education advisor is an expert in medical education and assessment. The advisor attends each MCCQE Part I test committee meeting, educates item writers, instructs members on the Blueprint and MCC Objectives, supports the TDOs in identifying content gap areas, and serves as a subject-matter expert across all test committees.

MCCQE Part I content is based primarily on topics that reflect the *MCC Objectives* and align with the approved MCCQE Blueprint. Item writers focus on specific Dimension of Care and Physician Activity topics from the Blueprint, based on gaps identified in the item bank. They are also asked to consider certain test specifications, such as gender, age group, and special populations, during question development as delineated in Table 2.

Each MCQ and CDM test committee reviews and approves new content for piloting. New questions are piloted and should their performance meet the statistical and content criteria, they would be counted as an active item and used in scoring.

2.3.1 Test committees

Each test committee has 8 to 10 subject matter experts from across Canada who have an interest and expertise in the fields of medical education and assessment. Each test committee consists of a minimum of two family physicians. Membership also includes representation from both official languages (English and French) as content is produced and/or translated in both official languages.

Test committee membership recommendations can come from TDOs, test committee members, or a member of MCC's selection committee. The selection committee reviews and approves appointment recommendations at the MCC's annual meeting and formally invites new members to be part of the recommended test committee.

Each test committee meets for two to three days at least once a year. During these meetings, MCQ and CDM items are written, classified, peer-reviewed and approved by the committee for piloting. There are additional quality assurance (QA) processes after the initial committee approval, including editorial, which is outlined below.

Committees develop content by following professional standards outlined in *The Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014), as well as the guidelines outlined in the *International Test Commission Guidelines on Test Use* (International Test Commission, 2001). These standards and guidelines include QA steps to ensure a fair assessment is delivered to the test takers.

Committee chairs and TDOs guide test committee members in the development of content where identified gaps exist in the exam Blueprint, and test specifications. Item development focuses on creating items that vary in difficulty and using the most up-to-date medical terminology (for example, compliant with the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders [DSM-5]* or newly established guidelines). Committee members focus the development of item content using specific in-practice examples along with anticipating where errors may occur.

After the test committee vets and approves items, the English content undergoes a rigorous editorial process by the English editors that includes substantive editing, copy editing, and proofreading. Substantive editing includes work to improve language and structure so that content is inclusive, clear, complete, and logical. Copy editing includes fact-checking and work to correct grammar, spelling, punctuation, and mechanics while ensuring consistency and adherence to style guidelines.

Since the MCC requires the highest quality of medical translation, all translators go through a screening process to evaluate their qualifications. After translation, a team of in-house French editors performs an in-depth comparative read to ensure that the translation is faithful to the English. This involves a thorough editorial and peer review process, in compliance with current French standards and MCC style guidelines. Once the edited content is approved by the TDOs and outstanding issues are resolved, the MCC conducts translation validation sessions with the French editors and francophone physicians to make any final correction or editorial content change and ensure that the French is inclusive of regional differences. The TDOs and editors then proofread all content for a final quality control review, before adding it to the pool of items available for selection and test assembly.

2.3.2 Clinical decision-making questions

The CDM test committee develops content for the CDM portion of the MCCQE Part I. This committee includes subject matter experts from across specialty areas (medicine; obstetrics

and gynecology; pediatrics; population health, ethics and legal organization of medicine [PHELO]; psychiatry; surgery; and family medicine). The CDM test committee has physician representation from both official languages (English and French). Gender diversity and geographic representation from across Canada are also a consideration in the committee membership. Similar to the content development of MCQs, the CDM test committee develops content by following professional standards mentioned in *Test committees*, section 2.3.1, and rigorous QA processes. Committee members meet twice a year, and their mandate is to create, review and classify CDM content based on existing Blueprint gaps.

The basis for the development of a CDM question is the key-feature approach. This approach is based on the notion of case specificity, which means that clinical performance on one problem may not be a good predictor of performance on other problems. Consequently, assessments of clinical performance need to sample broadly as skills do not generalize across problems. To sample broadly in a three-and-a-half-hour exam, it is important to focus on the key features in the resolution of each problem, be they essential issues or specific difficulties. Test committee members think about where a minimally competent candidate would likely make an error and use this as the focus for the development of key features.

The development of key feature–based cases for CDM has been guided by considerations of content validity, test score reliability, and sound principles of test development. Key feature cases provide flexibility in terms of question format (short-menu and write-in), multiple responses to items, and scoring criteria. Key feature problems have been found to be useful in assessments that require medical knowledge and the ability to apply that knowledge in clinical scenarios. These scenarios often require critical decisions to be made during the assessment and management of a given clinical problem. These specific, critical decision points constitute the key features of the problem.

Once test committee members have created and approved key features, they continue with case development. At this point, the test committee develops the case and questions in accordance with the scenario and the selected MCC Objective. The CDM scoring key reflects the main tasks that candidates must perform, which are identified in the key features. The CDM test committee approves all developed cases before they are piloted. As an additional QA step, the six MCQ specialty test committees vet the content. If necessary, they send feedback and suggest revisions to the CDM test committee. MPAC also reviews all CDM cases for final medical proofing. Once a case has been piloted and has performed well, the case is banked as an active case ready to be used on a future exam.

Item performance varies, and at times, items are flagged for psychometric reasons. Flagged items are reviewed prior to scoring the exam. Depending on the item, some content will be removed from scoring and must be sent back to the CDM Test Committee for review.

2.4 TEST ASSEMBLY

Following question development and piloting, fixed linear test forms are created to meet content specifications, test constraints and psychometric specifications. The number of forms is based on an analysis of the item bank. Due to the number of items per test form and the number of forms, computer software is used in the assembly of the test forms to ensure the construction of equivalent forms, both in content and in difficulty.

As part of test assembly, we also take into account the linking between test forms. Scores from different test forms are statistically linked through common items referred to as *anchor items*. Anchor items are assembled as a set of MCQs called *anchor sets*. Anchor items are selected using the content specifications to be a smaller representation of a complete exam in terms of both content and psychometric specifications and content constraints.

TDOs collaborate with psychometricians and physicians in the assembly of multiple test forms to ensure candidates receive a broad representation of content in their test-taking experience that is in line with the content specifications, test constraints and psychometric specifications. Other guidelines used in the assembly of the tests include ensuring the appropriate representation of topics of medicine, confirmation that items do not provide answers to other test questions and that item enemies (items of similar content) are tracked to avoid appearing on the same test form.

The TDOs and psychometricians work closely to ensure the test forms are reviewed and approved by subject matter experts. Once MCC staff has vetted the forms to ensure they meet the exam specifications, a committee of experts convenes once a year to review and approve the test forms. The first step is approval of the anchor items and the second is approval of the full test form of MCQs and CDMs. This is done by the test form approval committee (TFAC), which follows a thorough process to approve the test forms using the MCC's test form management system. The process for form approval is as follows:

1. The psychometrics and assessment services (PAS) staff assemble test forms according to the exam specifications.
2. The TDOs approve the forms, exchanging any items that overlap in content or that may be an item enemy and is not yet tagged in MCC's item bank. TDOs also identify any content that may be medically inaccurate (for example, if there have been any guideline changes since item development).
3. The TFAC approves the MCQ anchor sets first, as they establish the linking scale that connects all forms to ensure a comparable level of difficulty and precision. Once approved, the anchor sets are considered locked and cannot be replaced during the approval of an entire form.

4. The TFAC then reviews the remaining items on each test form and approves all the forms in their entirety.
5. A final review by PAS and the TDOs ensures the content specifications and constraints have been respected and the psychometric parameters are maintained in the final approved forms.

The MCCQE Part I has evolved from a semiadaptive exam, where questions candidates saw depended on their responses to previous items, to fixed examination forms where a preselected set of items is included in each form. MCC has developed automated methods for assembling test forms through constrained optimization that can most efficiently support the construction of multiple parallel test forms. After forms are assembled, they are reviewed and approved by the MCC's MCCQE Part I team (which includes item and test development experts and psychometricians) and two independent committees of physicians. Automated test assembly is used to assemble all MCCQE Part I test forms. Test forms are assembled to meet a series of content specifications, as described in Exam Specifications, section 2.2, and to be as similar as possible, both in content and in difficulty. Figure 2 depicts the logic implemented to automatically assemble a number of test forms. Common items are required to establish a common scale between different test forms. The result is that scores from different test forms can be compared as they share a common scale.

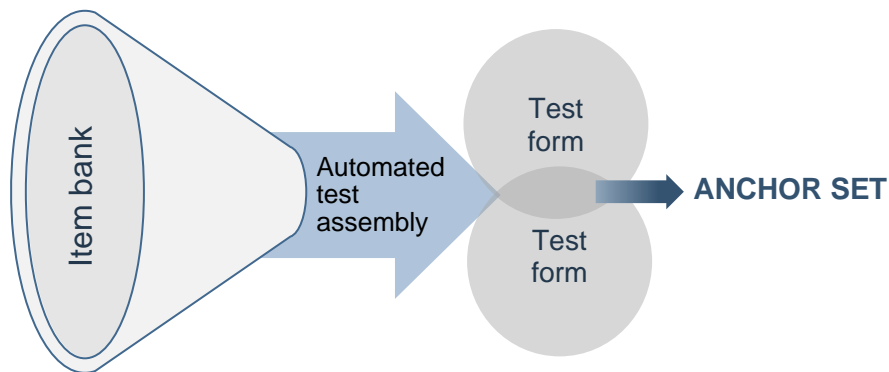


Figure 2: Automated test assembly procedure

3. Exam administration

3.1 EXAM DELIVERY

Starting in 2019, the MCCQE Part I was delivered in Canada and internationally in over 80 countries through Prometric, which has more than 20 years' experience in exam development and administration. Prometric is internationally recognized and serves professional high-stakes examination sectors. The change to Prometric ensures broader access for candidates to take the MCCQE Part I.

In 2020–2021, the MCCQE Part I was offered during four test windows in June, July, October, and January 2021. The test windows occurred over two- to nine-week period. The April session was postponed to June to meet candidate capacity demand in Canada, accommodate shifts in clerkship end dates, and account for test centre closures due to the pandemic. The postponement of the session also allowed for the exam delivery team to prepare to offer additional testing opportunities through remote proctoring. In both test centre and remote proctoring modalities, Prometric staff deliver the exam, follow strict security protocols, and monitor and support candidates' exam appointments from the registration stage through exam completion.

The exam may be taken in either English or French at any test centre and through remote proctoring; however, staff and technical support may provide service in only one language. In Canada, support in both official languages occurs at the Ottawa, Montreal, and Quebec City test centres.

3.2 EXAM SECURITY

The MCC takes several measures to safeguard exam security. In the COVID-19 context and with the introduction of remote proctoring exam security remains a priority.

Exam publishing processes are well established. Guidelines and security protocols are shared and reviewed for both test centre and remote proctoring administration before each testing window. Exam results are processed in MCC's secure environment. This cycle of exam delivery offers the MCC assurances of a consistent and fair exam administration for all candidates. The MCC collaborates with stakeholders on all facets of the exam process to ensure that only eligible candidates can write the exam and that no one has an unfair advantage.

Although remote proctoring poses new potential security risks (e.g., access to material that is not permitted, recording the exam), remote proctoring services are evolving and offering new AI-enabled features (e.g., facial detection) embedded in the testing software. These features are

integral to mitigate security risks. They capture eye movement or the presence of additional people in the testing room, block inappropriate keystrokes, and flag security violations such as the use of smartphones and other prohibited items. Other security features include lockdown browser functionality preventing candidates from accessing aids or capturing exam content and the use of video review if cheating is suspected.

Every site administrator and proctor is trained to recognize potential test security breaches. Training is standardized and delivered by Prometric across both modalities. Prometric conducts yearly training with all staff to communicate enhancements to security protocols and reinforce security measures. Prometric also performs regular test centre and remote proctoring audits to assess potential security gaps and ensure quick resolution. In addition to test security measures and a team that monitors exam activities throughout the examination session, MCC staff monitors online study forums for candidate activity around sharing of exam content before, during and after the administration.

Candidates taking an MCC examination have legal and professional responsibilities. The MCC also has a responsibility to candidates and to the Canadian population to ensure the integrity of its examinations. In 2018, the MCC introduced, as part of its registration and exam day process, an *Exam Test Security video*. All candidates need to agree to the terms and conditions, which state that they have understood the rules and regulations around test security. The creation of the video was in response to increased content breaches and a pattern from candidates that they were unaware that sharing exam content was in violation of their terms and conditions.

If a candidate appears to be giving or receiving information during the exam, the Prometric staff may immediately terminate the candidate's exam. The staff is required to produce a full candidate procedure report of all such occurrences to the MCC. The MCC also receives all candidate interaction logs to assess candidate behaviour and corroborate Prometric security concerns. All MCCQE Part I materials, including the content and questions comprising the MCCQE Part I, are protected by copyright and are to be kept confidential. Candidates are permitted to use the MCCQE Part I materials solely for the purpose of completing the MCCQE Part I and must not disseminate, reproduce, share or reveal to others the exam materials and content, in whole or in part, at any time or in any way, even after the exam ends. Comparing exam content and question themes with colleagues, sharing content with future exam candidates, and posting content online are considered breaches of confidentiality. Any breach of the MCCQE Part I Terms and Conditions is considered irregular behaviour for which the MCC or CEC may take appropriate action in accordance with the MCCQE Part I Terms and Conditions candidates accepted at the time of application. In the past, the CEC has issued a Denied Standing to candidates due to irregular behaviour; consequences of this can include the following:

- Candidates may be banned from taking future MCC examinations

- Candidates' physiciansapply.ca account may be suspended
- A permanent annotation may be made on candidates' physiciansapply.ca account
- A report may be made to medical regulatory authorities and other organizations
- Legal action may be taken against the candidate

3.3 EXAM PREPARATION

Online materials are available to assist candidates to prepare for the MCCQE Part I. These resources include the exam platform demonstration videos, sample MCQ and CDM questions, instructional videos (e.g., CDM tips, online demo), a list of resources by medical specialty area, and the MCC Objectives. All candidates have access to these *free online materials* through the MCC's website. A *self-education program* for physicians to learn about communication and cultural competencies required in Canada is offered through online modules on physiciansapply.ca.

Candidates may also test their knowledge by purchasing a full-length Preparatory Examination, shorter Preparatory Examination-Lite, and Multiple-Choice Questions and Clinical Decision-Making practice tests on the *MCC website*.

3.4 QUALITY ASSURANCE

After each exam day administration, MCC's database is updated with each candidate's response file. An initial system validation is done to ensure an accurate and complete candidate response file is received.

A second validation is completed at the end of the session. A table that includes one row per item for each candidate is generated for each exam component. The tables contain the unique identifiers for candidates and items along with the candidate answers and scores for all items. An initial round of quality assurance (QA) of the tables is performed by the psychometrician for the MCCQE Part I, including verification of completeness. Reasons for missing data are verified with the exams team. Once it is determined that the data meets the QA requirements, scoring and calibration are performed by Psychometric Assessment Services (PAS).

3.5 RELEASE OF RESULTS

Examination results are confirmed by the Central Examination Committee (CEC). Approximately seven weeks after the last day of the examination session, the CEC meets to review performance

on the exam, is informed about administrative issues, rules on special candidate cases, and approves exam results.

The MCC releases candidates' results (e.g., pass or fail decision) and their total score through their physiciansapply.ca account. Shortly thereafter, candidates have access to their Statement of Results (Appendix A), the official results document, and the Supplemental Information Report (Appendix B) that provides them with information on their strengths and weaknesses by the domains in the Blueprint.

4. Validity

It is generally accepted that tests are not inherently valid or invalid but that validity should be viewed as a process of gathering evidence that supports the intended interpretations and uses of test scores (AERA, APA, & NCME, 2014). Michael T. Kane (1990, 2013a, 2013b) has proposed an argument-based approach to validation that involves a process of gathering evidence to support score interpretations by establishing arguments that can be backed by theory, empirical research or common sense (Kane, 1990).

4.1 THE ARGUMENT-BASED APPROACH TO VALIDATION

According to Kane (2013b), the validity of a proposed interpretation and use depends on the plausibility of the claims being made, and validation involves the evaluation of these claims. Any claim that certain statements about score interpretations or uses being valid must be justified. Justification takes on the form of arguments. "Proposed interpretations and uses are valid to the extent that the reasoning involved in the interpretation is sound, reasonable, and plausible, that is, valid" (Kane, 1990).

For the MCCQE Part I, this entails gathering evidence to support the intended interpretations and uses of the examination. This means that scores and pass or fail decisions can be used to make valid decisions regarding the level of competence of a graduating student entering supervised practice. Validity considerations have been incorporated into exam design, exam specifications, item development, exam assembly, psychometric quality, exam administration and results reporting.

In Kane's approach, validating the interpretive arguments involves four inferences:

1. **Scoring:** Assigning scores to performance.
2. **Generalization:** From statements about observed performance to statements about expected performance over a universe of possible performances.
3. **Extrapolation:** Statements are extended to the expected performance over the domain.
4. **Implication:** Performance can also be used to make decisions about an examinee's future.

Figure 3 depicts Kane's framework for an argument-based approach to validation. His approach begins with an assessment of the scoring of a single observation such as responses to exam items (Scoring), to using the observed scores to generate an overall test score representing performance in the test setting (Generalization), to drawing an inference regarding what the test score might imply for real-life performance (Extrapolation), and finally to interpreting this information and making a decision (Implication).

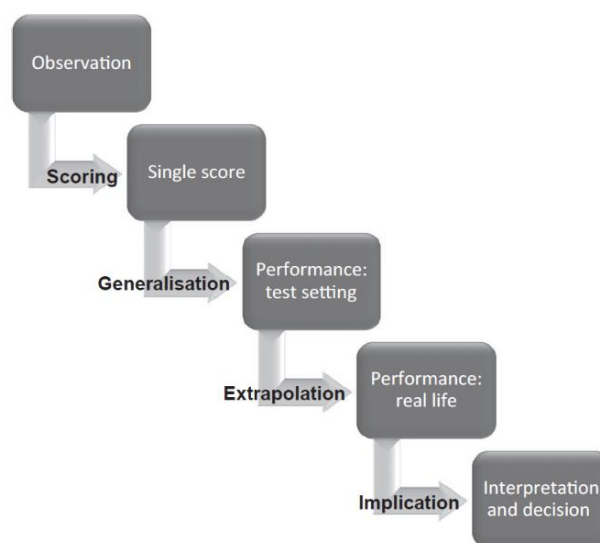


Figure 3: Key elements in Kane's argument-based approach to validation: Inferences from observation to decision (Source: Cook et al., 2015, p. 564)

In Tables 3 to 6, we provide evidence for the four levels of inference of Kane's argument-based approach to validation. In each of these tables, we present information about the Sources of Evidence (e.g., content expertise, test content, internal structure), Data (data used to support the claim), Warrant (logical statements that serve as bridges between the claim and the data), and Backing (additional justification for the warrant).

Table 3: Level of inference – Scoring

Sources of evidence	Data	Warrant	Backing
Based on content expertise	Documentation, meeting notes, training slides	Items are developed to reflect relevant medical ability	<ul style="list-style-type: none"> • During the course of exam content development, great care is taken to ensure the exam is relevant to medical graduates entering postgraduate training in Canada • As indicated in Exam Development, Section 2, items are developed based on the exam Blueprint and content specifications defined by the CEC members • CEC members ensure that the exam assesses the critical medical knowledge and clinical decision-making ability of a candidate at a level expected of a medical student who is completing their medical degree in Canada
Based on content expertise	Documentation, meeting notes, training slides	Proper training is offered for test developers	<ul style="list-style-type: none"> • Various test committees are involved in developing test items • Regular content development workshops are conducted to train test committee members to develop items that reflect the knowledge and skills emphasized in the content specifications and that meet professional test development guidelines • Guidelines have been developed for both MCQs and CDMs • The items are reviewed, edited and finalized by test committee members, test development officers (TDOs), editors, and translators. The items are reviewed, edited and finalized by test committee members, TDOs, editors, and translators
Based on content expertise	Documentation, meeting notes, training slides	Construct-irrelevant variance is minimized during item development	<ul style="list-style-type: none"> • During development, items are reviewed by subject matter experts (SMEs) and TDOs to ensure they meet the content specifications. • As well, SMEs, TDOs and editors review items for appropriateness of language and biased or noninclusive language or content • In addition, empirical evidence from the item and distractor analysis is used to further investigate potential sources of construct-irrelevant variance
Based on test content	Item responses and scoring	The answer keys are the	<ul style="list-style-type: none"> • Empirical evidence from item and distractor analyses are used to further investigate whether

Sources of evidence	Data	Warrant	Backing
	rules (MCQs and CDMs)	correct answers	<p>the answer key is the correct answer.</p> <ul style="list-style-type: none"> For example, item-total correlations for correct answers are positive and are negative for distractors Items not meeting this expectation are identified and provided to TDOs for content review before final calibration and test scoring
Evidence of precision	Write-in item responses	Markers are marking write-in responses consistently within an exam session	<ul style="list-style-type: none"> Each item is marked independently by two physician markers and when discrepancies are detected, the issue is resolved by a third marker CDM write-in items that display less than 90 per cent agreement between markers are flagged for review Additionally, items that have weighted kappa coefficients less than 0.61 are also flagged for review
Evidence of comparability	Candidate performance by delivery mode – 2018 to 2020	Average total scores	<ul style="list-style-type: none"> Candidate average total scores in 2020 (M=247) are fairly comparable with scores obtained in 2019 (M=252) and 2018 (M=250)
Evidence of comparability	Candidate performance by delivery mode	Average total scores	<ul style="list-style-type: none"> As data from the 2020 sessions shows, Canadian medical graduates (CMGs) taking the exam for the first time in a test centre had an average performance of (M=266), which is not significantly different from CMGs who took the exam remotely (M=268). For international medical graduates (IMGs) taking the exam for the first time, even though the difference between test centre (M=232) and remote proctoring (M=238) is statistically significant, the difference on the average total score is merely meaningful (approximately half of standard error of measurement and one fifth of standard deviation)
Evidence of comparability	Candidate performance by delivery mode	Pass rate	<ul style="list-style-type: none"> As data from the 2020 sessions shows, the pass rate for CMGs taking the exam for the first time in a test centre (98%) was exactly the same for CMGs taking the exam remotely (98%) For IMGs taking the exam for the first time, the difference between test centre (61%) and remote proctoring (69%) is statistically significant; however, confounding factors may be interfering in these results

Sources of evidence	Data	Warrant	Backing
			<ul style="list-style-type: none"> For example, confounding factors included the timing of the reopening of test centres and the registrations occurring in waves; CMGs were invited to register first (hence, reducing the spot availability for IMGs in test centres) Also, data analyses have indicated that more prepared CMG candidates registered to the initial exam dates This could have happened with IMG candidates as well; another hypothesis is that extra preparation could have affected borderline IMG candidates (the impact in pass rate is more pronounced than on test scores) The average total score for IMGs is close to the cut score (226), so small increases on their scores could cause a change in status from fail to pass

Table 4: Level of inference – Generalization

Sources of evidence	Data	Warrant	Backing
Evidence of precision	Item and test scores	The reported scores attain a level of decision accuracy and decision consistency that meets the target values	<ul style="list-style-type: none"> Using data from 2019 and 2020, the decision consistency estimates have varied from 0.86 to 0.96 and the decision accuracy estimates from 0.90 to 0.97, which indicates reliable and valid pass or fail decisions Values were above the target value of 0.80
Evidence of precision	Item and test scores	The reported scores attain the level of precision required for a high-stakes exam; total score reliability estimates are above the target value of 0.80	<ul style="list-style-type: none"> Considering the 2019 and 2020 sessions, the person [test] reliability estimates have varied from 0.86 to 0.93, indicating an adequate level of reliability of test scores, given the characteristics of the population of examinees (i.e., high achievers)
Based on test content	Blueprint classification	Test forms are comparable in	<ul style="list-style-type: none"> Automated test assembly (ATA) was used to assemble a number of fixed linear test forms, meeting almost perfectly the content

Sources of evidence	Data	Warrant	Backing
		content	specifications, as described in Exam Development, Section 2
Based on test content	Item parameters	Test forms are comparable in levels of difficulty	<ul style="list-style-type: none"> • During ATA, test forms were assembled to also be as similar in difficulty as possible • The test information function for each of the test forms was inspected and results support the parallelism among the different test forms
Based on test internal structure	Correlation between domains and total score	Blueprint domains are highly correlated with total score	<ul style="list-style-type: none"> • Correlations from spring 2018 suggest that the MCCQE Part I measures an essentially single dominant underlying construct (i.e., basic medical knowledge and clinical skills that the MCCQE Part I is designed to measure) • All domains were found to be significantly, positively correlated with one another (see Appendix C) • The highest correlation was found with the total score • Correlations were also computed using the raw scores and results support the same conclusion • This provides preliminary evidence to support the assumption of unidimensionality underlying the use of the Rasch model used to assemble and score the exam

Table 5: Level of inference – Extrapolation

Sources of evidence	Data	Warrant	Backing
Evidence of precision	Item and test scores	The correlation between the MCCQE Part I and NAC Examination provide some evidence of convergent validity	<ul style="list-style-type: none"> • The relationships between scores on the MCCQE Part I and the NAC Examination were also investigated • The NAC Examination uses an objective structured clinical examination (OSCE) format to assess the readiness of an IMG for entry into a Canadian residency program • A significant correlation ($r = .55, p < .0001$) was obtained between scores on the MCCQE Part I and the NAC Examination based on a sample of 87 candidates whose scores on both exams were matched using data from spring 2018 • The correlation is strong enough to provide some evidence of convergent validity between the two

Sources of evidence	Data	Warrant	Backing
			<p>MCC exams, but not too high to indicate redundancy, as the two exams are assessing different aspects of clinical knowledge and skills</p> <ul style="list-style-type: none"> • Caution advised in interpreting this result due to low number of candidates taking both exams (N=87)

Table 6: Level of inference – Decisions

Sources of evidence	Data	Warrant	Backing
Based on standard setting	MCCQE Part I test scores and pass or fail status; subject matter expertise	Those who pass the MCCQE Part I are competent enough to practise medicine safely and efficiently	<ul style="list-style-type: none"> • The cut score is reflective of a point on the proficiency scale that represents the minimum standard • After a comprehensive standard-setting procedure with 22 panellists, the MCC's CEC endorsed a pass score of 226 on a scale of 100 to 400 as a defensible standard to apply starting with the April 2018 administration • Evidence of validity indicating that MCCQE Part I meets best practices when setting new pass scores includes <ul style="list-style-type: none"> ○ careful selection of panellists ○ careful training of panellists ○ the methodology used on the standard-setting exercise followed best practices (Bookmark and Hofstee methods) ○ feedback of the panellists post standard-setting exercise • Internal evidence included the consistency of the panellists and convergence of results • Two subpanels arrived at a similar pass score independently at 95% confidence intervals constructed using standard error of judgment (SEJ) • SEJ indicates the variability that would be expected if the same judging process were repeated by many different panels of similar composition • More information on the standard-setting procedure can be found here

5. Psychometric analyses

In 2020–2021, the MCCQE Part I was offered during four test sessions in June 2020, July 2020, October 2020 and January 2021.

In this section, we describe the psychometric analyses completed following the June exam session. We conduct item analyses, followed by item calibration, estimation of candidates' ability, scoring, standard setting and scaling (when applicable), and finally, score reporting.

5.1 ITEM ANALYSIS: CLASSICAL TEST THEORY AND ITEM RESPONSE THEORY

Following the June session, a comprehensive set of item analyses was conducted to verify the soundness of each item from a statistical perspective before engaging in final scoring of the exam. Item analysis, using both classical test theory and item response theory, results in items being flagged for various reasons outlined below. The inclusion or exclusion of items flagged during item analysis in final scoring is predicated on a careful content review by experts. While content experts are encouraged to use the statistical information in the review process, the final decision rests on whether the content is defensible given the intent of the item and/or case.

Classical test theory and item response theory flags

Immediately following a session, an initial item analysis (IIA) is conducted using responses from all Canadian medical graduates taking the exam for the first time. An IIA involves a classical item analysis to review item difficulty, discrimination, and candidate raw-score performance. Specifically, p-values are computed as a measure of an item's difficulty and an item-measure correlation is computed to reflect item discrimination.

In addition, psychometric and assessment services (PAS) examines the proportion of candidates who select each option as an indicator of how well each distractor (an incorrect response) is functioning. The investigation of how well each distractor is performing is supported by computing the correlation between each distractor and the total score. If distractors are performing as intended, these correlations will be negative (for example, candidates with lower overall MCCQE Part I scores are selecting the distractors more frequently than higher-ability candidates). Furthermore, items with near zero option endorsement (for example, too few candidates who obtain a particular score or choose a particular distractor) are also flagged for content review.

Since the adoption of the partial credit model (Masters, 1982), a model of the family of Rasch models, for the calibration and scoring in the spring 2015 MCCQE Part I, additional statistical criteria have been introduced for the CDM component to identify potentially flawed items.

Currently, the CDM component has dichotomous as well as polytomous items. For polytomous items, the partial credit model is used to establish the difficulty level for the transitions (i.e., steps) between successive item scores. These transitions are modelled using step parameters (or step thresholds), and are expected to increase in value as the score categories increase. It is expected that candidates' average abilities advance across categories for CDM items. That is, a score of 2 on an item requires higher overall ability than a score of 1. When this expectation is not met, these items are referred to as having disordered step parameters (for instance, weaker candidates overall on the exam obtain higher scores on the item than more able candidates). These items are flagged as potentially flawed and subject to content review.

Additionally, CDM write-in items that display less than 90 per cent agreement between markers or have a weighted kappa coefficient of less than 0.61 are also flagged for review. The kappa coefficient reflects the agreement between markers beyond chance agreement (Cohen, 1979), as it is expected that scores assigned by two markers would yield highly comparable results.

Items flagged by PAS are reviewed by both psychometricians and content experts. An item is flagged if it meets one or more of the following rules:

- Very high difficulty: $p\text{-value} < 0.10$
- Very low difficulty: $p\text{-value} > 0.95$
- High percentage of omits: > 5 per cent
- Low correlation value for the correct answer: < 0.05
- High correlation value for distractor: > 0.05 and $N > 10$
- Top 20 per cent performers chose distractor more often than the correct answer
- Item mean square outfit < 0.5
- Item mean square outfit > 2.0 .
- Low category score frequency $N < 10$
- Disordered threshold (CDM only)
- Average ability not increasing (CDM only)
- Per cent agreement < 0.90 (write-in only)
- Weighted kappa < 0.61 (write-in only)

Flagged items are included in final item response theory (IRT) calibrations only after psychometricians and content experts have reviewed the items and confirmed that the content is

acceptable and the key is correct. Items flagged during initial item analysis and determined to be flawed after review are removed from further analyses with the review committee's approval. The fall sessions are processed using the same item difficulty estimates from spring so that scores are on the same scale and thus comparable.

5.2 ITEM CALIBRATION

Previous research studies (De Champlain et al., 2016; Morin et al., 2014) have established that simpler models such as the Rasch model yield results that are consistent with those from more elaborate models such as the two-parameter IRT logistic model. Starting with the spring 2015 administration, the Rasch model and one of its extensions, the partial credit model (Masters, 1982), were applied using Winsteps (Linacre, 2015) to the MCCQE Part I for item calibration and scoring. This transition has allowed the implementation of a unified IRT model for the estimation of all MCQ and CDM dichotomous and polytomous items as well as establishing candidate abilities by considering all items together (MCQs and CDMs).

With the Rasch model for dichotomous data, the probability of a correct response on an item is modelled as a logistic function of the difference between the ability of a person and the item difficulty parameter. If $X = 1$ denotes a correct response and $X = 0$ denotes an incorrect response, the probability of a correct response takes on the following form:

$$P_i\{X_{ni}\} = \frac{e^{\beta_n - \delta_i}}{1 + e^{\beta_n - \delta_i}}$$

where β_n is the ability of person n and δ_i is the difficulty of item i .

For polytomous items, the partial-credit model is a generalization of the dichotomous model. It is a general measurement model that provides a theoretical foundation for the use of sequential integer scores (categorical scores).

For the 2020 MCCQE Part I, items were recalibrated maintaining the scale established in 2018. Data from first-time Canadian medical graduate test takers was used for this process. First, the parameters for all the active items were estimated to identify potential poor-performing items. Through this step, items that did not satisfy the statistical criteria outlined in Item Analysis: Classical Test Theory and Item Response Theory, Section 5.1, were flagged and reviewed by subject matter experts (SMEs). The decision to retain or remove those items from scoring was made. After SMEs review all flagged items (in step 1) and decide which items to remove from scoring and calibration, items are recalibrated excluding those items. A final set of calibrated items are then ready to use in estimating candidates' abilities.

5.3 ESTIMATING CANDIDATE ABILITY

After items are vetted by SMEs, item parameters are used for estimating the ability of all candidates. Item parameters are fixed in the estimation process and only the level of candidate ability is estimated.

A candidate's ability and total score on the MCCQE Part I is derived from their combined performance on the MCQ and CDM components. The MCC uses the partial credit model (Masters, 1982) to score candidates' exam responses. While raw score data (scores of 1 point or zero points) are necessary, they are insufficient to establish a candidate's ability level. Simply adding up item scores does not accurately reflect a candidate's ability since this does not take into account the difficulty level of the items that were encountered in any given MCCQE Part I form.

MCQ and CDM short-menu items are machine-scored as they involve numbered responses that are then compared to predefined scoring keys. To ensure correctness in the scoring process, a rigorous QA process is implemented at this stage: test items are independently scored (using the predefined scoring keys) by two statistical analysts using two different types of statistical software. Results are compared and after a 100% match, they are reviewed by the psychometrician to ensure reasonableness.

CDM write-in items are marked by physician markers. Since the fall 2014 MCCQE Part I, physician markers have used the MCC-developed software to facilitate the marking of CDM constructed response items. Physician markers are presented with CDM cases, items, key features and scoring keys. Before the answers are presented, the software combines identical answers given by candidates for a given item. All unique answers that do not aggregate are also presented. Physician markers are then asked to indicate whether an answer is deemed correct or incorrect given predetermined scoring keys. Each item is marked independently by two physician markers; if discrepancies are detected, the issue is resolved by a third marker.

The software also allows physician markers to indicate whether candidates have exceeded the number of answers allowed for an item. Markers do not assign scores to items; they are simply asked to indicate whether answers are correct or incorrect and scoring is performed following this validation step. Once all answers have been categorized as either correct or incorrect, scoring is done automatically, taking into account all other constraints such as exceeding the maximum number of answers allowed. The process of attributing scores to the CDM write-in items is similar to the MCQ and CDM short-menu items described above. In other words, it goes through the same rigorous QA process.

All MCQs are dichotomously scored as they all have one correct answer. Sometimes, CDM items can also be dichotomously scored. For polytomous CDM items that involve more than one correct

answer, successive integer scores are assigned, also called category scores. For example, a candidate selecting two out of three correct answers would receive two points.

The Rasch model also allows us to establish a scale that is expressed in such a way that candidate attributes, such as ability, and item attributes, such as item difficulty, are on the same unit of measurement. In its initial phase, a scale is defined in measurement units called logits (log-odds units) and allows for candidates' abilities to be expressed on the same scale as the item difficulties. Values typically range between -3.0 and $+3.0$, although values beyond the latter can occur. A candidate who obtains a score of -3.0 would demonstrate very little ability in regard to the specialty areas being assessed, whereas a candidate who obtains a score of $+3.0$ would demonstrate strong ability.

5.4 STANDARD SETTING AND SCALING

The MCC conducts a standard-setting exercise every three to five years to ensure the standard and the pass score remain appropriate. Standard setting is a process used to define an acceptable level of performance and to establish a pass score.

In the summer of 2018, the MCC completed a rigorous standard-setting exercise¹ based on expert judgments from a panel of 22 physicians representing faculties of medicine from across the country, different specialties and years of experience supervising students and residents. The Bookmark Method, a successfully employed and defended method used by large-scale exam programs, was used to help panellists suggest a new pass score for the exam. The recommended pass score was subsequently brought forward to the CEC for consideration and approval. The CEC, whose members are appointed annually by the MCC's Council, was responsible for the quality of MCC examinations and awards final results, such as pass or fail, to candidates. The CEC approved the recommended pass score.

In the spring 2018 MCCQE Part I, a new pass score was applied to reflect this minimally acceptable level of performance. The value representing this standard was established at 0.682 on the logit scale. Though the logit scale defined above has properties that are well suited for mathematical calculations, it is not very user-friendly for the candidate population. A linear transformation of the ability estimate is necessary to establish a scale of reported scores that is more meaningful to candidates. The scale chosen has a mean of 250 and a standard deviation of

¹ See [*MCCQE Part I 2018 Standard setting report*](#).

30 based on all first-time candidates in spring 2018. On that scale, the pass score is equivalent to 226 for the MCCQE Part I.

To establish an individual candidate's scale score, a linear transformation is performed. The following generic formula is applied:

$$X'_i = a + bX_i$$

Where X'_i = scaled score;

a = *the additive component often referred to as the intercept;*

b = the multiplicative component of the linear transformation
often referred to as the slope;

And X_i = a candidate's Rasch ability score

In the spring of 2018, when the scale was first established, the slope was 58.46300753 and the intercept was 185.7324343. These two constants were applied to transform each candidate's ability score, estimated using the partial credit model, into a scale score.

A candidate's final result such as pass or fail is determined by their total score and where it falls in relation to the exam pass score; a total score equal to or greater than the pass score is a pass and a total score less than the pass score is a fail. The candidate's performance is judged in relation to the exam pass score and not judged on how well other individuals perform.

5.5 SCORE REPORTING

Approximately seven weeks after the last day of the exam session, the MCC issues a Statement of Results (SOR) and a Supplemental Information Report (SIR) to each candidate through their physiciansapply.ca account. A sample of the SOR is included in Appendix A and a sample of the SIR is in Appendix B. The SOR includes the candidate's final result and total score as well as the score required to pass the exam. Additional information about subscores and comparative information is provided in the SIR, offering the candidate information on areas of strengths and weaknesses. Since subscores have fewer items, there is less measurement precision. Subscores are provided to individual candidates graphically and for feedback only and are not meant to be used by organizations for selection purposes.

If a candidate's performance has potentially been affected by procedural irregularities that occurred during an exam, it is reported to the CEC for a special ruling. A candidate may receive a No Standing, as the MCC cannot, in these cases, establish a valid pass or fail decision. In other special cases, such as candidates having been observed violating the exam's regulations (for example, having been observed using a smartphone during the exam), the MCC may award a Denied Standing.

6. Exam results

Candidate performance for the four sessions in 2020–2021 is summarized in this section. When applicable, historical data from previous years are included for reference.

6.1 CANDIDATE COHORTS

The 2020–2021 MCCQE Part I includes data from June 2020, July 2020, October 2020 and January 2021 sessions. In 2020, the exam was administered in a seven-week session (June 1 to July 19), a nine-week session (July 20 to September 20) and a two-week session (October 26 to November 8). In 2021, it was administered in a four-week session (January 25 to February 21). A total of 6,731 candidates took the exam across 143 countries. In 2020–2021, two candidates received a No Standing or Denied Standing. Table 7 summarizes the distribution of candidates across groups defined by their country of graduation and whether they were a first-time or repeat test taker of the MCCQE Part I.

Table 7: MCCQE Part I group composition, 2020–2021
(percentages do not always total 100 due to rounding)

Group	Jun. 2020		Jul. 2020		Oct. 2020		Jan. 2021	
	No. of test takers	% of total	No. of test takers	% of total	No. of test takers	% of total	No. of test takers	% of total
Canadian medical graduate (CMG) first-time test takers	2,245	84.0	476	25.4	111	8.4	74	8.5
CMG repeat test takers	17	0.6	31	1.7	19	1.4	19	2.2
International medical graduate (IMG) first-time test takers	304	11.4	982	52.4	875	66.5	551	63.4
IMG repeat test takers	106	4.0	386	20.6	310	23.6	225	25.9
Total	2,672	100	1,875	100	1,315	100	869	100

6.2 OVERALL EXAM RESULTS

Table 8 shows pass rates and basic statistics. On the score reporting scale of 100 to 400, the pass score is 226. This table does not include the two candidates with No Standing or Denied Standing.

Table 8: MCCQE Part I results, 2020–2021

		Jun. 2020	Jul. 2020	Oct. 2020	Jan. 2021
Canadian medical graduate (CMG) first-time test takers	No. of test takers	2,245	476	111	74
	Mean score	269	260	262	268
	Standard deviation	20.0	20.8	24.0	19.9
	Min. score	197	196	202	229
	Max. score	343	323	330	314
	Pass rate (%)	98.6	95.8	94.6	100
CMG repeat test takers	No. of test takers	17	31	19	19
	Mean score	244	245	239	241
	Standard deviation	12.5	18.2	13.7	14.7
	Min. score	220	210	214	217
	Max. score	267	272	263	275
	Pass rate (%)	94.1	80.6	89.5	84.2
International medical graduate (IMG) first-time test takers	No. of test takers	304	981	875	551
	Mean score	243	235	231	232
	Standard deviation	32.8	35.8	36.1	37.2
	Min. score	118	100	100	100
	Max. score	315	315	311	315
	Pass rate (%)	70.4	64.8	61.5	60.8
IMG repeat test takers	No. of test takers	105	386	310	225
	Mean score	228	224	221	223
	Standard deviation	23.6	25.3	25.9	25.1
	Min. score	157	136	100	100
	Max. score	271	281	279	288
	Pass rate (%)	60.0	51.3	50.0	52.4
All candidates	No. of test takers	2,671	1,874	1,315	869
	Mean score	264	239	232	233
	Standard deviation	24.6	33.0	34.4	34.7
	Min. score	118	100	100	100
	Max. score	343	323	330	315
	Pass rate (%)	93.9	70.2	62.0	62.5

Figure 4 displays the total score distribution on the reported score scale for all candidates in the four sessions and total. Overall, the total score performance of the June cohort was better than the other three cohorts.

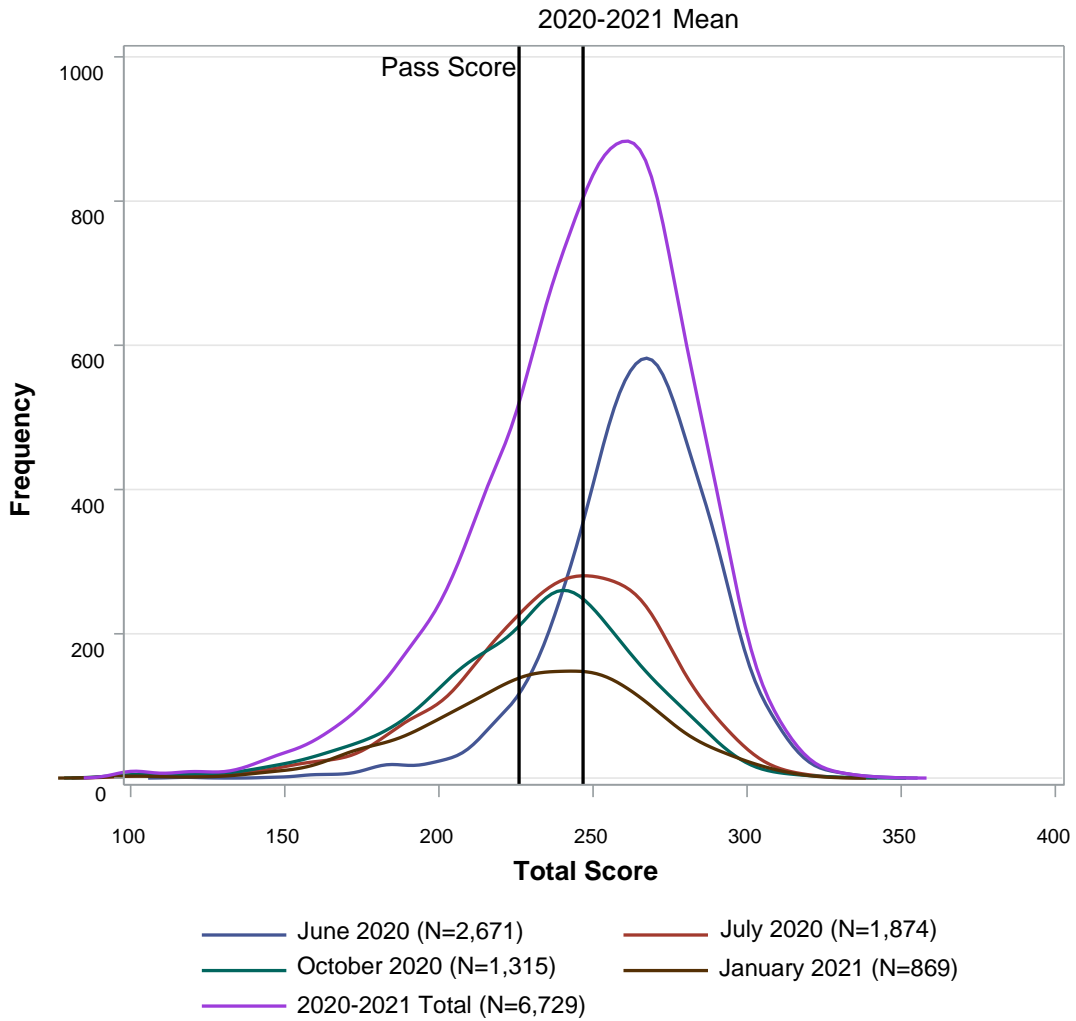


Figure 4: MCCQE Part I total score distributions, 2020–2021

6.3 RELIABILITY OF EXAM SCORES AND CLASSIFICATION DECISIONS

Test reliability refers to the extent to which the sample of items in any exam accurately measures the intended construct. Reliability of the MCCQE Part I can be assessed by examining the standard error (SE) of ability estimates along the reported score scale. The SE indicates the precision with which the scores are reported at a given point on the scale and is inversely related to the amount of information provided by a test at that point. The SE values should be as small as possible so that the measurement of the candidate’s ability contains as little error as possible. In the framework of item response theory (IRT), the SE serves the same purpose as the standard error of measurement (SEM) in classical measurement theory (Hambleton et al., 1991), except that the estimation of SE varies with ability levels in IRT whereas the most common estimation methods of the classical SEM do not.

Figures 5 through 8 display scatter plots of SE values along the reported score scale for the four 2020–2021 sessions. For each cohort, the plot shows that scores are less accurate toward the lower and higher ends of the score scale, but more accurate in the middle range of the scale where the majority of the scores fall. The SE is lower near the pass score, which indicates highest precision of ability estimates, thus supporting more accurate and consistent pass or fail decisions.

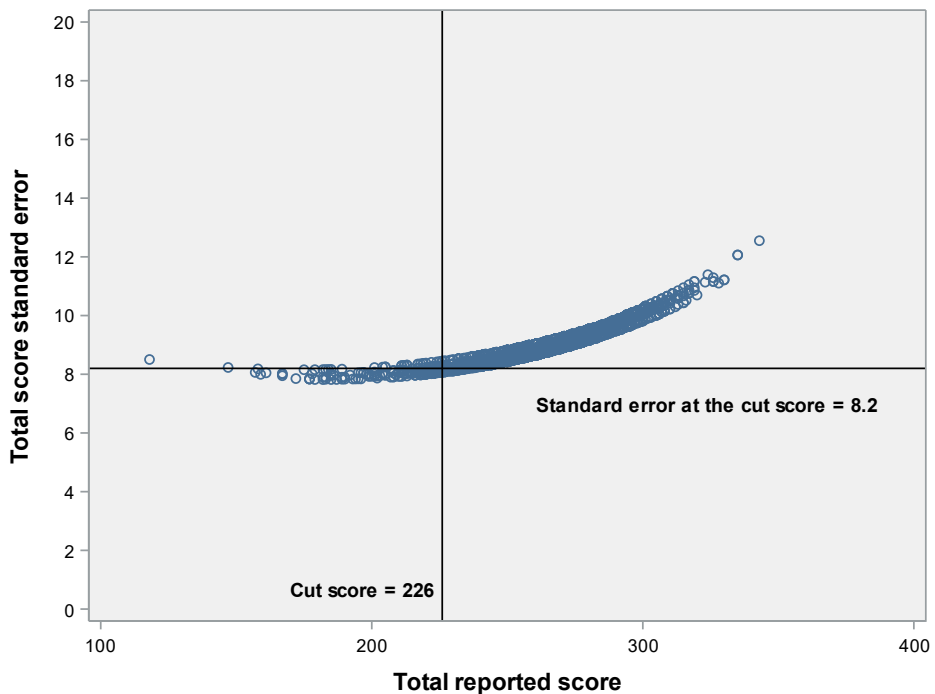


Figure 5: MCCQE Part I standard error of ability estimates, June 2020

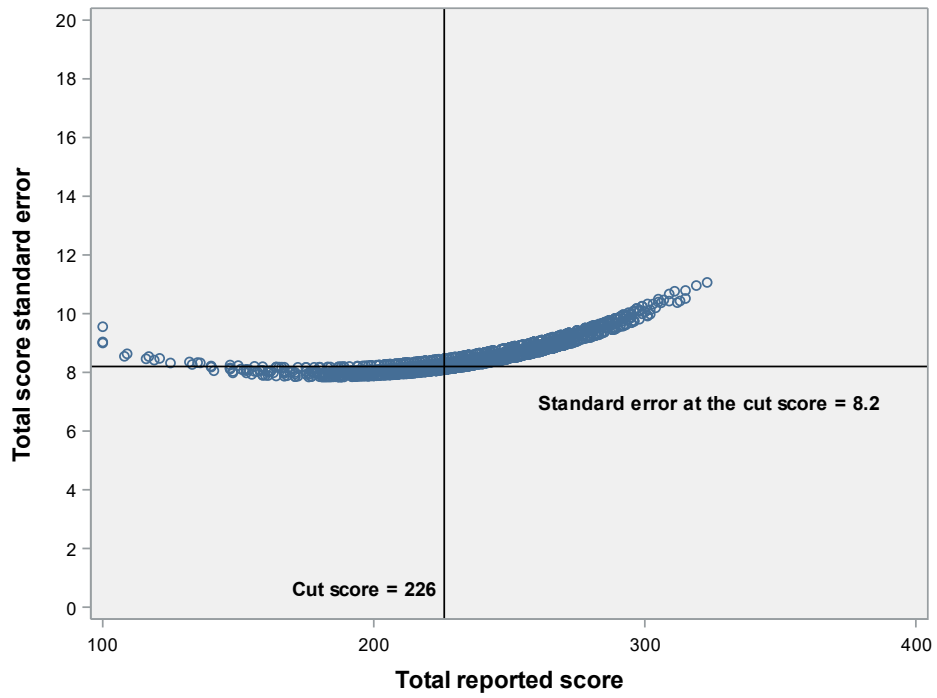


Figure 6: MCCQE Part I standard error of ability estimates, July 2020

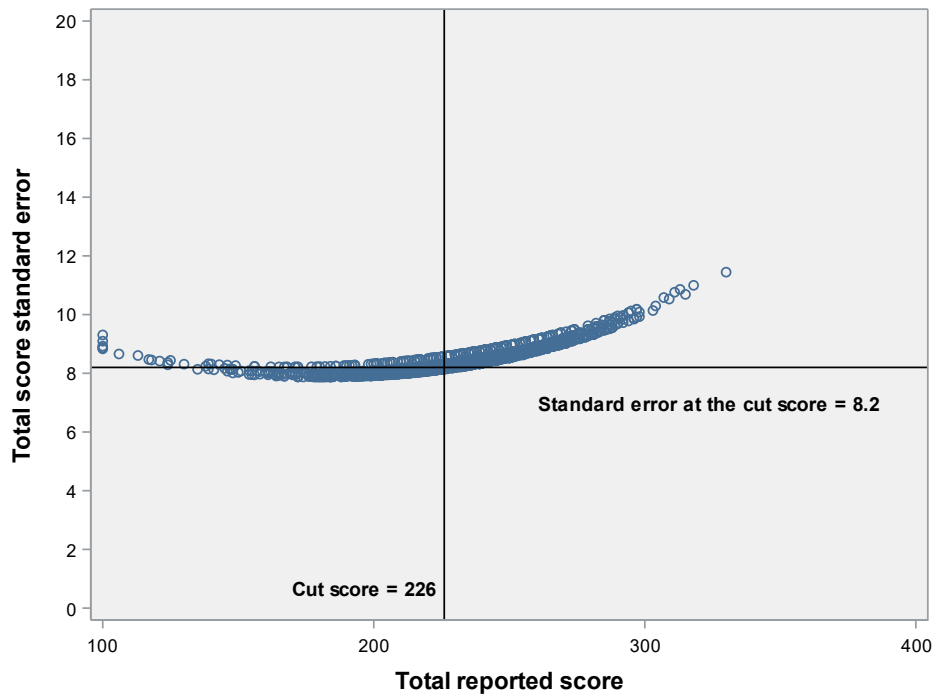


Figure 7: MCCQE Part I standard error of ability estimates, October 2020

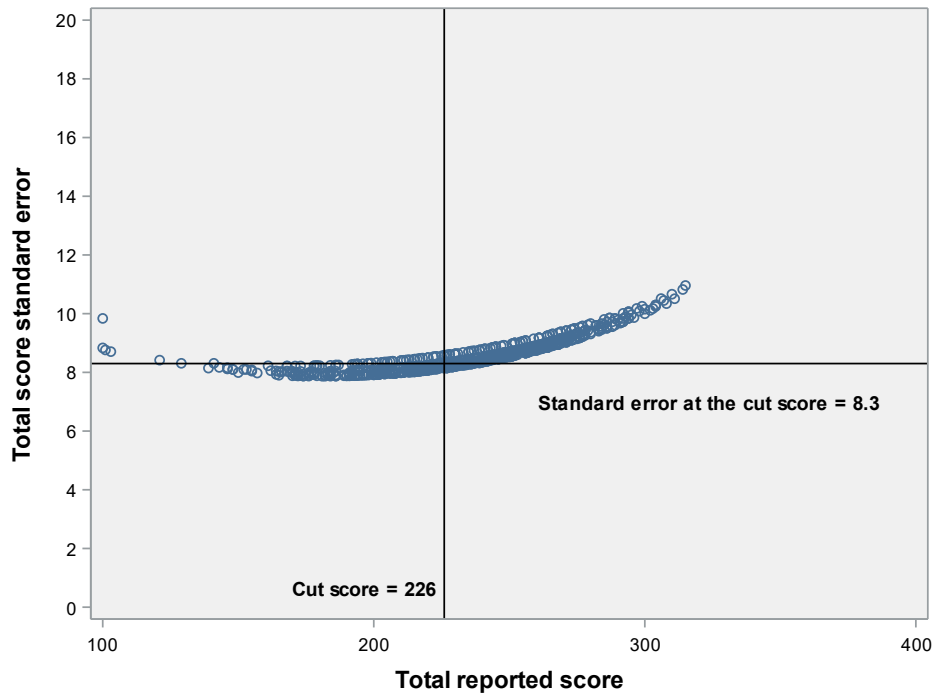


Figure 8: MCCQE Part I standard error of ability estimates, January 2021

6.4 PASS OR FAIL DECISION ACCURACY AND CONSISTENCY

In the context of this high-stakes exam, the accuracy of pass or fail decisions is of the utmost importance. Decision consistency and decision accuracy can be estimated using the Livingston and Lewis procedure (Livingston & Lewis, 1995) that is used by many high-stakes testing programs. Decision consistency is an estimate of the agreement between pass or fail final decisions on potential parallel forms of the exam. Decision accuracy is the estimate of the agreement between the pass or fail decisions based on observed exam scores and those that would be based on their true score (for example, if the candidate could be tested on an infinite number of MCCQE Part I items). As indicated in Table 9, both the decision consistency estimate and the decision accuracy estimate for each of the four sessions of 2020–2021 indicate reliable and valid pass or fail decisions based on MCCQE Part I scores. Table 9 is based on data from 2,671 candidates in the June 2020 session, 1,874 in the July 2020 session, 1,315 in the October 2020 session, and 869 in the January 2021 session.

Table 9: Reliability estimates, standard errors of measurement, decision consistency and decision accuracy indices for each MCCQE Part I session, 2020–2021

	June	July	Oct.	Jan.
Reliability estimate¹	0.86	0.92	0.93	0.93
Average standard error of measurement (SEM) (total score)	9.0	8.6	8.5	8.6
Decision consistency	0.96	0.90	0.89	0.89
False positive	0.02	0.05	0.05	0.06
False negative	0.02	0.05	0.05	0.05
Decision accuracy	0.97	0.93	0.92	0.92
False positive	0.01	0.03	0.04	0.04
False negative	0.02	0.04	0.04	0.04

¹ Person (test) reliability from the Rasch model.

6.5 DOMAIN SUBSCORE PROFILE

The purpose of the domain subscore profile is to provide diagnostic information to candidates by highlighting their relative strengths and weaknesses. The Supplemental Information Report (SIR) is designed to provide subscore information at the candidate level.

Domain subscore information for all candidates in the 2020–2021 sessions is provided below. The range of domain subscores is presented graphically in Figures 9 to 12. The graphs show the domain subscore for each of the eight domains. The boxes for each domain indicate the range of scores for 50 per cent of the candidates' domain subscores. The vertical line represents the median or 50th percentile subscore. The remaining 50 per cent of domain subscores are shown to the right or the left of the box as a line (25 per cent to the right and 25 per cent to the left).

The legend for each of the subscores displayed in Figures 9 to 12 is as follows:

Dimensions of Care	Physician Activities
HEALTHP = Health promotion and illness prevention	PSYCHS = Psychosocial aspects
ACUTE = Acute	MGMT = Management
CHRONIC = Chronic	COMM = Communication
PSYCHS = Psychosocial aspects	PROFB = Professional behaviours

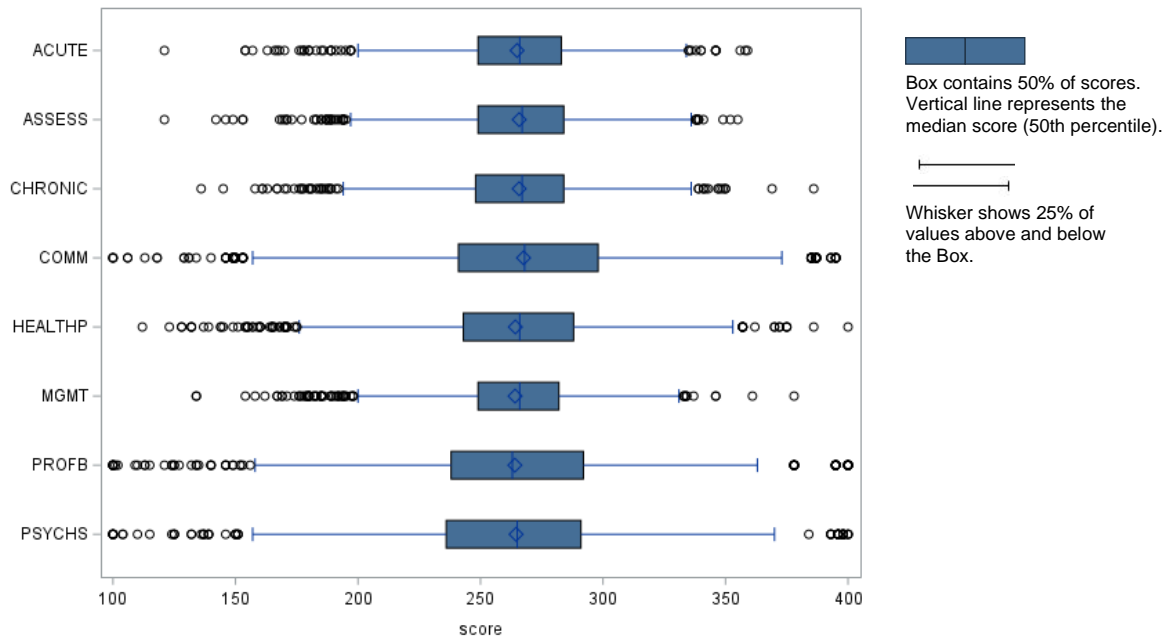


Figure 9: Domain subscore for the MCCQE Part I, June 2020 session

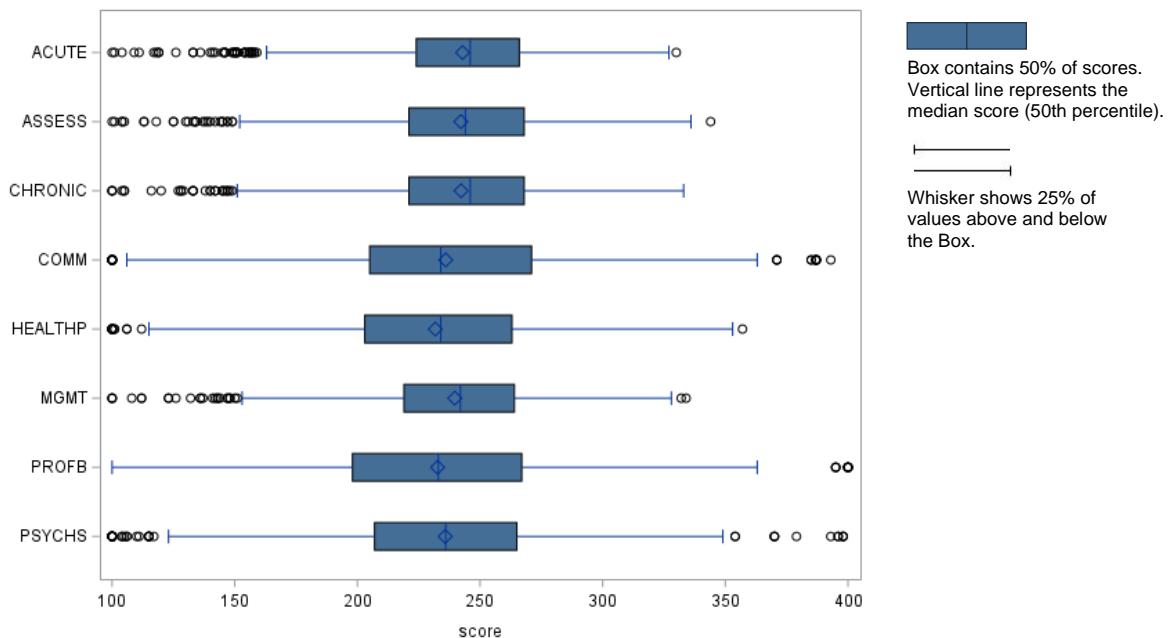


Figure 10: Domain subscore for the MCCQE Part I, July 2020 session

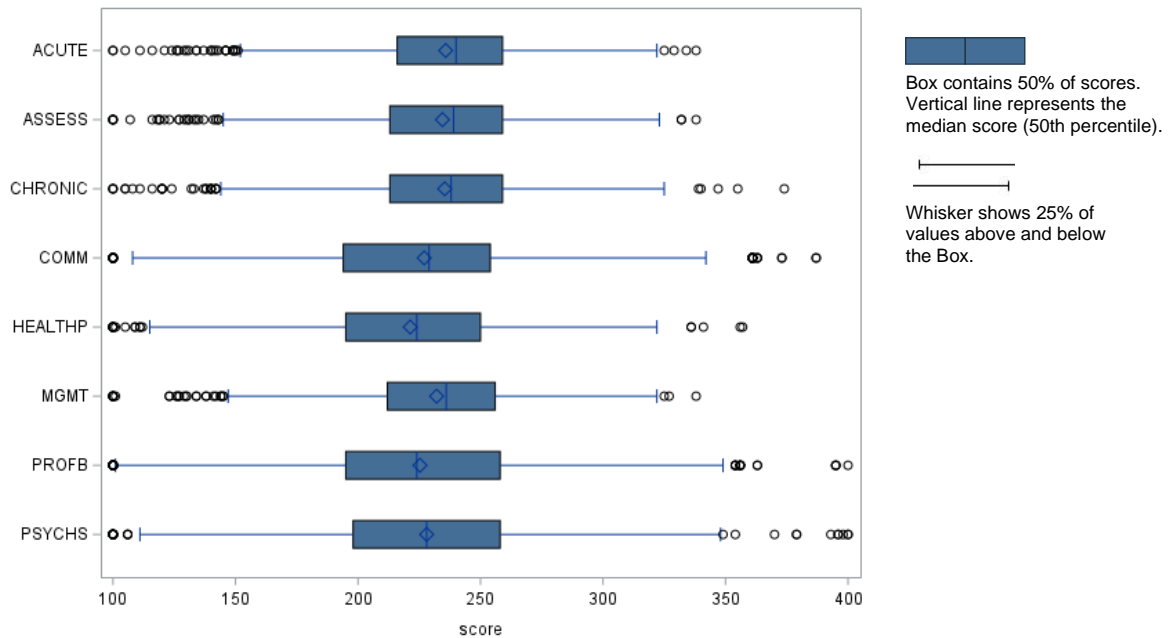


Figure 11: Domain subscore for the MCCQE Part I, October 2020 session

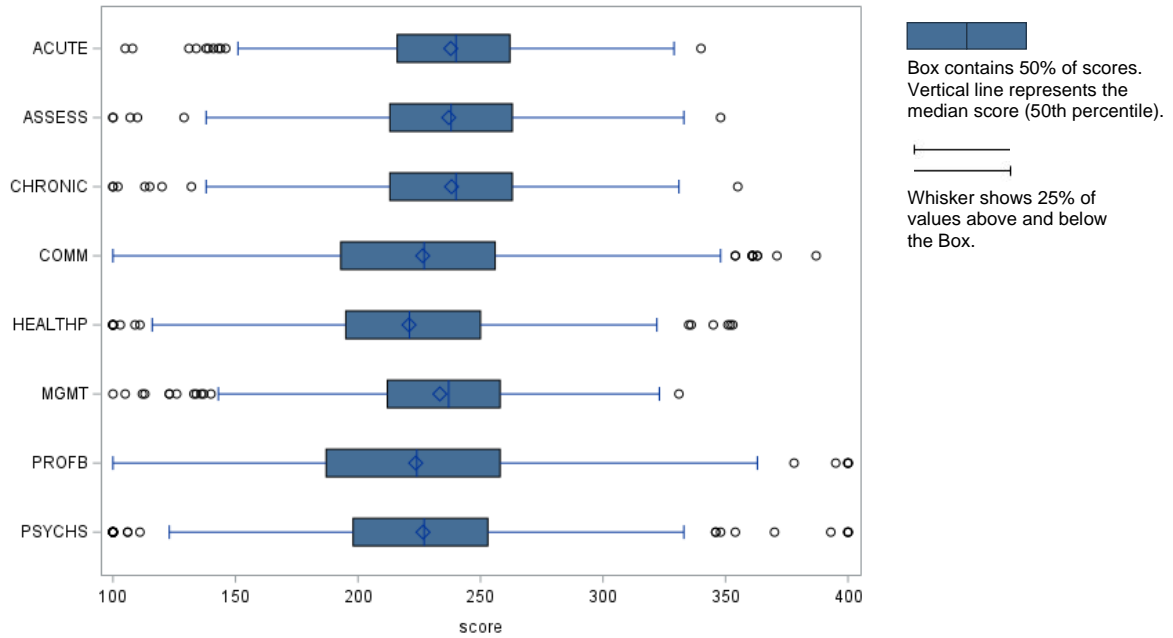


Figure 12: Domain subscore for the MCCQE Part I, January 2021 session

6.6 HISTORICAL PASS RATES

Historical pass rates are presented in this section. Table 10 shows the pass rates for 2017 to 2020–2021 by group.

Table 10: MCCQE Part I pass rates, April 2017 to January 2021

	2017		2018		2019–2020		2020–2021	
	No. of test takers	Pass rate (%)	No. of test takers	Pass rate (%)	No. of test takers	Pass rate (%)	No. of test takers	Pass rate (%)
Canadian medical graduate (CMG) first-time test takers	2,802	95	2,823	95	2,861	97	2,906	98
CMG repeat takers	156	63	178	67	138	73	86	86
International medical graduate (IMG) first-time test takers	1,677	62	1,413	62	3,929	53	2,711	64
IMG repeat takers	1,264	29	991	24	950	42	1,026	52
TOTAL	5,899	71	5,405	73	7,878	68	6,729	77

6.7 CANDIDATE SURVEY

A survey is administered with each exam to gather candidate feedback regarding their test-taking experience. Tables 11 and 12 present the results of the survey questions, post-MCQ session and post-CDM session, respectively, to which candidates responded. In Tables 11 and 12, the percentage in the Missing column is equal to the number of missing answers divided by the sum of the number of missing answers and valid answers.

Table 11: MCCQE Part I post-MCQ candidate survey results, 2020–2021
Percentages for ratings

The MCQ section provided an opportunity for me to demonstrate my level of medical knowledge and clinical decision-making before entering supervised practice.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	Total Valid	Missing
Jun. 2020	100 (4%)	340 (13%)	710 (28%)	1,316 (51%)	104 (4%)	2,570 (100%)	102 (4%)
Jul. 2020	42 (2%)	157 (9%)	369 (21%)	946 (54%)	231 (13%)	1,745 (100%)	130 (7%)
Oct. 2020	29 (2%)	78 (6%)	256 (21%)	684 (56%)	181 (15%)	1,228 (100%)	87 (7%)
Jan. 2021	20 (2%)	45 (6%)	148 (18%)	438 (55%)	150 (19%)	801 (100%)	68 (8%)

The questions were clearly written.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	Total Valid	Missing
Jun. 2020	57 (2%)	376 (15%)	639 (25%)	1,335 (52%)	163 (6%)	2,570 (100%)	102 (4%)
Jul. 2020	41 (2%)	164 (9%)	368 (21%)	930 (54%)	235 (14%)	1,738 (100%)	137 (7%)
Oct. 2020	23 (2%)	127 (10%)	219 (18%)	659 (53%)	204 (17%)	1,232 (100%)	83 (6%)
Jan. 2021	13 (2%)	58 (7%)	138 (17%)	451 (56%)	141 (18%)	801 (100%)	68 (8%)

How would you rate the performance/stability of the computer application so far during your exam?

	Poor	Fair	Good	Very Good	Excellent	Total Valid	Missing
Jun. 2020	324 (13%)	376 (15%)	651 (25%)	679 (26%)	535 (21%)	2,565 (100%)	107 (4%)
Jul. 2020	83 (5%)	163 (9%)	366 (21%)	555 (32%)	575 (33%)	1,742 (100%)	133 (7%)
Oct. 2020	57 (5%)	100 (8%)	291 (24%)	359 (29%)	419 (34%)	1,226 (100%)	89 (7%)
Jan. 2021	27 (3%)	56 (7%)	176 (22%)	250 (31%)	294 (37%)	803 (100%)	66 (8%)

Overall, how would you rate the usability of the user interface?

	Poor	Fair	Good	Very Good	Excellent	Total Valid	Missing
Jun. 2020	76 (3%)	331 (13%)	881 (34%)	893 (35%)	379 (15%)	2,560 (100%)	112 (4%)
Jul. 2020	23 (1%)	128 (7%)	465 (27%)	675 (39%)	445 (26%)	1,736 (100%)	139 (7%)
Oct. 2020	19 (2%)	84 (7%)	342 (28%)	430 (35%)	343 (28%)	1,218 (100%)	97 (7%)
Jan. 2021	9 (1%)	38 (5%)	230 (29%)	292 (37%)	228 (29%)	797 (100%)	72 (8%)

Table 12: MCCQE Part I post-CDM candidate survey results, 2020–2021
Percentages for ratings

The tutorial at the beginning of the exam was helpful.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	Total valid	Missing
Jun. 2020	58 (2%)	178 (7%)	662 (26%)	1,366 (54%)	259 (10%)	2,523 (100%)	149 (6%)
Jul. 2020	22 (1%)	65 (4%)	283 (16%)	1,019 (59%)	342 (20%)	1,731 (100%)	144 (8%)
Oct. 2020	14 (1%)	37 (3%)	194 (16%)	681 (56%)	294 (24%)	1,220 (100%)	95 (7%)
Jan. 2021	7 (1%)	20 (3%)	130 (16%)	443 (56%)	192 (24%)	792 (100%)	77 (9%)

The CDM section provided an opportunity for me to demonstrate my level of medical knowledge before entering supervised practice.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	Total valid	Missing
Jun. 2020	150 (6%)	498 (20%)	808 (32%)	969 (38%)	100 (4%)	2,525 (100%)	147 (6%)
Jul. 2020	59 (3%)	235 (13%)	461 (26%)	811 (47%)	176 (10%)	1,742 (100%)	133 (7%)
Oct. 2020	42 (3%)	161 (13%)	291 (24%)	565 (46%)	158 (13%)	1,217 (100%)	98 (7%)
Jan. 2021	25 (3%)	71 (9%)	188 (24%)	390 (50%)	113 (14%)	787 (100%)	82 (9%)

The CDM cases and corresponding questions were clearly written.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	Total valid	Missing
Jun. 2020	166 (7%)	680 (27%)	726 (29%)	872 (35%)	78 (3%)	2,522 (100%)	150 (6%)
Jul. 2020	75 (4%)	292 (17%)	456 (26%)	763 (44%)	150 (9%)	1,736 (100%)	139 (7%)
Oct. 2020	60 (5%)	178 (15%)	291 (24%)	540 (45%)	144 (12%)	1,213 (100%)	102 (8%)
Jan. 2021	28 (4%)	119 (15%)	176 (22%)	361 (46%)	101 (13%)	785 (100%)	84 (10%)

Was your exam delivered via a remote proctor, or did you take it in a Prometric test centre with an in-person proctor? (PS: self-reported question)

	Remote proctor	In-person proctor at a prometric test centre	Total valid	Missing
Jun. 2020	1,767 (70%)	751 (30%)	2,518 (100%)	154 (6%)
Jul. 2020	567 (33%)	1,172 (67%)	1,739 (100%)	136 (7%)
Oct. 2020	459 (38%)	756 (62%)	1,215 (100%)	100 (8%)
Jan. 2021	286 (36%)	499 (64%)	785 (100%)	84 (10%)

How would you rate the performance/stability of the computer application during your exam?

	Poor	Fair	Good	Very Good	Excellent	Total Valid	Missing
Jun. 2020	350 (14%)	380 (15%)	631 (25%)	669 (27%)	490 (19%)	2,520 (100%)	152 (6%)
Jul. 2020	115 (7%)	125 (7%)	348 (20%)	632 (36%)	517 (30%)	1,737 (100%)	138 (7%)
Oct. 2020	66 (5%)	99 (8%)	246 (20%)	424 (35%)	377 (31%)	1,212 (100%)	103 (8%)
Jan. 2021	38 (5%)	50 (6%)	181 (23%)	273 (35%)	245 (31%)	787 (100%)	82 (9%)

Overall, how would you rate the usability of the user interface?

	Poor	Fair	Good	Very Good	Excellent	Total Valid	Missing
Jun. 2020	205 (8%)	430 (17%)	791 (31%)	767 (30%)	326 (13%)	2,519 (100%)	153 (6%)
Jul. 2020	66 (4%)	144 (8%)	457 (26%)	668 (39%)	399 (23%)	1,734 (100%)	141 (8%)
Oct. 2020	51 (4%)	117 (10%)	313 (26%)	414 (34%)	317 (26%)	1,212 (100%)	103 (8%)
Jan. 2021	16 (2%)	73 (9%)	203 (26%)	293 (38%)	196 (25%)	781 (100%)	88 (10%)

How would you rate your proctor's helpfulness?

	Poor	Fair	Good	Very Good	Excellent	Total Valid	Missing
Jun. 2020	122 (5%)	199 (8%)	677 (27%)	852 (34%)	661 (26%)	2,511 (100%)	161 (6%)
Jul. 2020	25 (1%)	104 (6%)	402 (23%)	628 (37%)	554 (32%)	1,713 (100%)	162 (9%)
Oct. 2020	11 (1%)	57 (5%)	279 (23%)	410 (34%)	441 (37%)	1,198 (100%)	117 (9%)
Jan. 2021	14 (2%)	37 (5%)	180 (23%)	288 (37%)	256 (33%)	775 (100%)	94 (11%)

The testing environment was conducive to writing a high-stakes exam (e.g., the noise level was appropriate and there were no major disruptions or interruptions).

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	Total valid	Missing
Jun. 2020	141 (6%)	231 (9%)	482 (19%)	1169 (47%)	478 (19%)	2,501 (100%)	171 (6%)
Jul. 2020	46 (3%)	77 (4%)	250 (15%)	845 (49%)	503 (29%)	1,721 (100%)	154 (8%)
Oct. 2020	22 (2%)	73 (6%)	150 (13%)	595 (50%)	356 (30%)	1,196 (100%)	119 (9%)
Jan. 2021	22 (3%)	44 (6%)	87 (11%)	395 (51%)	223 (29%)	771 (100%)	98 (11%)

How would you rate your overall exam experience?


	Poor	Fair	Good	Very Good	Excellent	Total Valid	Missing
Jun. 2020	384 (15%)	623 (25%)	826 (33%)	526 (21%)	149 (6%)	2,508 (100%)	164 (6%)
Jul. 2020	93 (5%)	264 (15%)	634 (37%)	551 (32%)	182 (11%)	1,724 (100%)	151 (8%)
Oct. 2020	56 (5%)	175 (15%)	416 (35%)	375 (31%)	175 (15%)	1,197 (100%)	118 (9%)
Jan. 2021	28 (4%)	93 (12%)	281 (36%)	250 (32%)	121 (16%)	773 (100%)	96 (11%)

7. References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Cohen, L. (1979). Approximate expressions for parameter estimates in the Rasch model. *The British Journal of Mathematical and Statistical Psychology*, 32, 113–120. onlinelibrary.wiley.com/doi/10.1111/j.2044-8317.1979.tb00756.x/abstract
- Cook, D. A., Brydges R., Ginsburg S., & Hatala R. (2015). A contemporary approach to validity arguments: a practical guide to Kane's framework. *Med Educ.*, 49(6):560–75. doi.org/10.1111/medu.12678
- De Champlain, A. F., Boulais, A.-P., & Dallas, A. (2016). Calibrating the Medical Council of Canada's Qualifying Examination Part I using an integrated item response theory framework: a comparison of models and designs. *J Educ Eval Health Prof*, 13:6. doi.org/10.3352/jeehp.2016.13.6
- Frank, J. R., Snell, L., & Sherbino, J. (2015). *CanMEDS 2015 Physician Competency Framework*. Royal College of Physicians and Surgeons of Canada. Retrieved from researchgate.net/publication/289254803_CanMEDS_2015_Physician_Competency_Framework
- Gierl, M. J., & Haladyna, T. (2013). *Automatic item generation: Theory and practice*. New York: Routledge.
- Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of item response theory*. Sage.
- International Test Commission (2001). International guidelines for test use. *International Journal of Testing*, 1(2), 93–114.
- Kane, M. (1990). *An argument-based approach to validation*. American Coll. Testing Program.
- Kane, M. (2013a). The argument-based approach to validation. *School Psychology Review*, 42(4), 448–457.
- Kane, M. (2013b). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1–73.
- Linacre, J. M. (2015). *Winsteps* (Version 3.91.0) [Computer software]. Retrieved from winsteps.com

- Linacre, J. M. (2016). *Winsteps Rasch measurement computer program User's Guide*. Winsteps.com.
- Livingston, S. A., & Lewis C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement*, 32(2), 179–197. jstor.org/stable/1435147
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174. [dx.doi.org/ 10.1007/BF02296272](https://dx.doi.org/10.1007/BF02296272)
- Morin, M., Boulais, A-P., & De Champlain, A. (2014). *Scoring the Medical Council of Canada's Qualifying Examination Part I: A comparison of multiple IRT models using different calibration methods*. Unpublished paper.
- Muchinsky, P. M. (1996). The correction for attenuation. *Educational & Psychological Measurement* 56:1, 63–75.
- Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. The University of Chicago Press.

APPENDIX A: MCCQE Part I Statement of Results sample



MEDICAL COUNCIL OF CANADA LE CONSEIL MÉDICAL DU CANADA

Medical Council of Canada
Qualifying Examination Part I
Statement of Results

Candidate name:	Aaa-Aaaaa, Bbbbbb		
Candidate code:	1234567890		
Examination session:	October 2020	Your final result:	Pass
Pass score:	226	Your total score:	288

December 17, 2020

We are writing to inform you of your final result on the Medical Council of Canada Qualifying Examination Part I.

Your total score is reported as a scaled score ranging from 100 to 400 with a mean of 250 and a standard deviation of 30. The mean and standard deviation were set using the results from the April 2018 session.

Your final result is based on your total score relative to the pass score.

For more information, please visit the exam's Scoring web page on our website, mcc.ca.

Supplemental information on your examination performance is reported to you in a separate document within your physiciansapply.ca account.

mcc.ca
physiciansapply.ca
inscriptionmed.ca

APPENDIX B: MCCQE Part I Supplemental Information Report sample



Medical Council of Canada Qualifying Examination Part I Supplemental Information Report

Candidate name: Aaa-Aaaaa, Bbbbbb
Candidate code: 1234567890
Examination session: October 2020
Your final result: Pass
Your total score: 288

This report provides you with supplemental information on your performance on the Medical Council of Canada Qualifying Examination (MCCQE) Part I.

The MCCQE Part I assesses the critical medical knowledge and clinical decision-making ability of a candidate at a level expected of a medical student who is completing his or her medical degree in Canada.

The exam assessed your performance across two broad categories with each exam question classified on both categories:

- Dimensions of care, covering the spectrum of medical care;
- Physician activities, reflecting a physician’s scope of practice.

Each category has four domains:

Dimensions of Care	Physician Activities
Health Promotion and Illness Prevention	Assessment and Diagnosis
Acute Care	Management
Chronic Care	Communication
Psychosocial Aspects	Professional Behaviours

See p. 3 of this report for the domain definitions.

Figure 1 displays your performance in each domain under Dimensions of Care. Figure 2 displays your performance in each domain under Physician Activities.

In both figures, we provide your subscores along with the mean subscore of first-time takers who passed the same exam in spring 2018 when the reporting scale and pass score were established.

Each domain is assigned a weighting on the exam. We present the content weights, expressed as percentages, in the grids shown on page 3.

We also provide the standard error of measurement (SEM) for each of your subscores. It represents the expected variation in your subscore if you were to take this exam again with a different set of questions covering the same domains.

Small differences in subscores or overlap between SEMs indicate that performance in those domains was somewhat similar. Overlap between the SEM and the mean score of first-time takers who passed signifies that performance is similar to the mean.

Subscores are based on less data than the total score and have less precision.

For more information, please visit the exam’s Scoring web page on our website mcc.ca.

mcc.ca
physiciansapply.ca
inscriptionmed.ca

Figure 1: Dimensions of Care

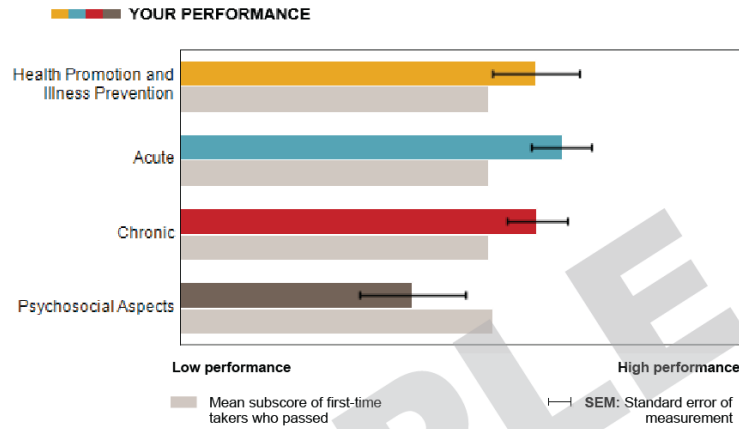
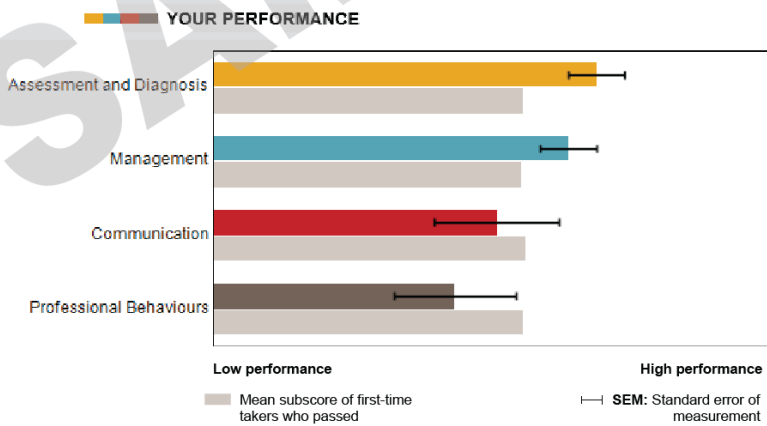


Figure 2: Physician Activities



Dimensions of Care

Reflects the focus of care for the patient, family, community and/or population:

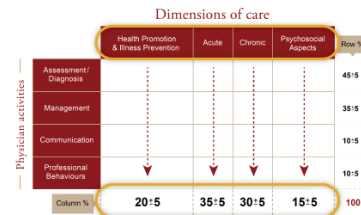
- **Health Promotion and Illness Prevention:**

The process of enabling people to increase control over their health and its determinants, and thereby improve their health. Illness prevention covers measures not only to prevent the occurrence of illness, such as risk factor reduction, but also to arrest its progress and reduce its consequences once established. This includes, but is not limited to screening, periodic health exam, health maintenance, patient education and advocacy, and community and population health.

- **Acute:** Brief episode of illness within the time span defined by initial presentation through to transition of care. This dimension includes but is not limited to urgent, emergent, and life-threatening conditions, new conditions, and exacerbation of underlying conditions.

- **Chronic:** Illness of long duration that includes but is not limited to illnesses with slow progression.

- **Psychosocial Aspects:** Presentations rooted in the social and psychological determinants of health and how these can impact on wellbeing or illness. The determinants include but are not limited to life challenges, income, culture, and the impact of the patient's social and physical environment.



Physician Activities

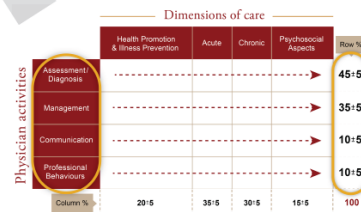
Reflects the scope of practice and behaviours of a physician practicing in Canada:

- **Assessment/Diagnosis:** Exploration of illness and disease using clinical judgment to gather, interpret and synthesize relevant information that includes but is not limited to history taking, physical examination and investigation.

- **Management:** Process that includes but is not limited to generating, planning, organizing safe and effective care in collaboration with patients, families, communities, populations, and other professionals (e.g., finding common ground, agreeing on problems and goals of care, time and resource management, roles to arrive at mutual decisions for treatment, working in teams).

- **Communication:** Interactions with patients, families, caregivers, other professionals, communities and populations. Elements include but are not limited to relationship development, intra-professional and inter-professional collaborative care, education, verbal communication (e.g., using the patient-centered interview and active listening), non-verbal and written communication, obtaining informed consent, and disclosure of patient safety incidents.

- **Professional Behaviours:** Attitudes, knowledge, and skills relating to clinical and/or medical administrative competence, communication, ethics, as well as societal and legal duties. The wise application of these behaviours demonstrates a commitment to excellence, respect, integrity, empathy, accountability and altruism within the Canadian health-care system. Professional behaviours also include but are not limited to self-awareness, reflection, life-long learning, leadership, scholarly habits and physician health for sustainable practice.



APPENDIX C: Internal structure of the MCCQE Part I

The Medical Council of Canada (MCC) undertook a strategic review of its assessment processes with a clear focus on their purposes and objectives and their structure and alignment with MCC's major stakeholder requirements. The review addressed current trends in medical education, regulation and assessment. The review also considered the role and purpose of the MCC's examinations in meeting the current and future needs of medical regulatory authorities (MRAs), the public and other stakeholders. In addition to focusing on the reassessment and realignment of the MCC's exams, a key recommendation focused on validating and updating the blueprints for both the multiple-choice question (MCQ) and clinical decision-making (CDM) components of the MCC Qualifying Examination (MCCQE) Part I.

As part of its commitment to adhere to best practices in medical education and assessment, the MCC undertook a Blueprint project to review and establish an evidence-based approach for identifying the competencies that physicians will be expected to demonstrate and be assessed on at two decision points: (1) entry into residency and (2) entry into independent practice. The purpose is to ensure that critical core competencies, knowledge, skills and behaviours for safe and effective patient care in Canada are being appropriately assessed for the two decision points. The rigorous and consultative process of how the Blueprint was developed can be found [here](#).

With the new Blueprint, the MCC has the opportunity to assess fundamental core competencies required of physicians practising in Canada at various points along their careers, regardless of specialties, and considers the performance across two broad categories, Dimensions of Care and Physician Activities. The internal structure of the MCCQE Part I can be revealed, to some degree, through the evaluation of the correlations among the Blueprint subscores. Correlating the two categories (and their embedded domains) can help in understanding how closely the exam conforms to the construct of interest. Correlations among subscores were examined using the data from 4,166 examinees who took the MCCQE Part I in the April 2018 session.

Table 13: Correlation matrix among subscores in the four domains of Dimensions of Care and total scores

	Total Score	Health Promotion	Acute	Chronic	Psychosocial Aspects
Total Score	1				
Health promotion and illness prevention	0.84	1			
Acute	0.91	0.66	1		
Chronic	0.86	0.64	0.68	1	
Psychosocial aspects	0.67	0.53	0.51	0.48	1

Table 14: Correlation matrix among subscores in the four domains of Physician Activities and total scores

	Total Score	Assessment and Diagnosis	Management	Communication	Professional Behaviours
Total Score	1				
Assessment and diagnosis	0.91	1			
Management	0.92	0.74	1		
Communication	0.67	0.50	0.55	1	
Professional behaviours	0.67	0.49	0.55	0.47	1

Table 15: Correlation matrix among subscores in Physician Activities and in Dimensions of Care

	Health Promotion and Illness Prevention	Acute	Chronic	Psychosocial Aspects
Assessment and diagnosis	0.72	0.87	0.81	0.52
Management	0.79	0.84	0.80	0.58
Communication	0.64	0.54	0.53	0.61
Professional behaviours	0.59	0.55	0.51	0.66

As indicated in Tables 13 to 15, all subscores classified by either Dimensions of Care or Physician Activities were found to be significantly and positively correlated with one another.

This provides preliminary evidence to support the assumption of unidimensionality underlying the use of the Rasch model used to assemble and score the exam. Correlations were also computed using the raw scores, and results support the same conclusion of unidimensionality.